

BATMAN: Bayesian Technique for Multi-image Analysis

J. Casado^{1,2,3*}, Y. Ascasibar^{1,2}, R. García-Benito⁴, G. Guidi³,
O. S. Choudhury³, E. Bellocchi^{1,2}, S. F. Sánchez⁵ and A.I. Díaz^{1,2}

¹ *Universidad Autónoma de Madrid, 28049 Madrid, Spain*

² *Astro-UAM, UAM, Unidad Asociada CSIC*

³ *Leibniz-Intitut für Astrophysik Potsdam (AIP), D-14482, Potsdam, Germany*

⁴ *Instituto de Astrofísica de Andalucía (IAA/CSIC), 18080 Granada, Spain*

⁵ *Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, 04519, México*

Accepted (20 December 2016)

ABSTRACT

This paper describes the Bayesian Technique for Multi-image Analysis (BATMAN), a novel image-segmentation technique based on Bayesian statistics that characterizes any astronomical dataset containing spatial information and performs a tessellation based on the measurements and errors provided as input. The algorithm iteratively merges spatial elements as long as they are statistically consistent with carrying the same information (i.e. identical signal within the errors). We illustrate its operation and performance with a set of test cases including both synthetic and real Integral-Field Spectroscopic data. The output segmentations adapt to the underlying spatial structure, regardless of its morphology and/or the statistical properties of the noise. The quality of the recovered signal represents an improvement with respect to the input, especially in regions with low signal-to-noise ratio. However, the algorithm may be sensitive to small-scale random fluctuations, and its performance in presence of spatial gradients is limited. Due to these effects, errors may be underestimated by as much as a factor of two. Our analysis reveals that the algorithm prioritizes conservation of all the statistically-significant information over noise reduction, and that the precise choice of the input data has a crucial impact on the results. Hence, the philosophy of BATMAN is not to be used as a ‘black box’ to improve the signal-to-noise ratio, but as a new approach to characterize spatially-resolved data prior to its analysis. The source code is publicly available at <http://astro.ft.uam.es/SELGIFS/BatMAN>.

Key words: methods: data analysis – methods: statistical – methods: numerical – techniques: image processing

1 INTRODUCTION

One of the basic problems in astronomical data analysis is the characterization of spatially-resolved information and the measurement of physical properties (and their variation) across extended sources. Soon this kind of tasks became increasingly demanding as new observations started to provide larger and larger amounts of data, prompting the need for a certain level of automation. Nowadays, human supervision is often limited to problematic cases, and many current and forthcoming datasets are so vast that a significant part of the analysis is left entirely to computer programs.

One of the first instances of such application is the identification of (potentially extended and/or blended) sources

in photometric images. Since the advent of large extragalactic surveys in the late 70’s, a wide variety of techniques have been developed in order to automatically create source catalogues from astronomical images (e.g. Stetson 1987; Bertin & Arnouts 1996; Makovoz & Marleau 2005; Savage & Oliver 2007; Molinari et al. 2011; Hancock et al. 2012; Men’shchikov et al. 2012, among many others). The widespread use of Integral-Field Spectroscopy (IFS) has literally added a new dimension to the problem, in the sense that spatial and spectral information are combined in order to locate the sources (see e.g. Koribalski 2012, and references therein, in the context of radio observations of the 21-cm line).

Recently, new algorithms working on different types of spatially-resolved data have been developed in order to tackle specific scientific problems beyond source detection. In many cases, the aim is to compute and characterize maps

* E-mail: javier.casado@uam.es

that trace the spatial distribution of a given physical quantity such as e.g. the temperature and composition of the hot intracluster medium (Sanders & Fabian 2001; Diehl & Statler 2006), the properties of the stellar population in early-type galaxies (Cappellari & Copin 2003), the moments of the velocity distribution along the line of sight (Krajinović et al. 2006), or the emission of warm ionized gas and HII regions in galaxies (Papaderos et al. 2002; Sánchez et al. 2012b, 2016).

What most of these goals have in common is the challenge of performing a coherent spatial segmentation of an image (or an IFS datacube) as a first step. On the one hand, it is necessary to average the signal over a large region in order to increase the signal-to-noise ratio (hereafter S/N) and carry out meaningful measurements. On the other hand, averaging over too large area does not only lead to a loss in spatial resolution, but it may also (in some cases, strongly) bias the results and their physical interpretation if the signal within the chosen region is not homogeneous.

As mentioned above, this problem has already been tackled by different authors. They have proposed several schemes, specifically optimized for a wide variety of problems, to divide an image into connected regions in a fully automatic way. Some of them (e.g. Sanders & Fabian 2001; Cappellari & Copin 2003; Diehl & Statler 2006) are based on a Voronoi tessellation, whereas some others rely upon the identification of suitable intensity thresholds (e.g. Stetson 1987; Bertin & Arnouts 1996; Sánchez et al. 2012b) or isocontours (e.g. Papaderos et al. 2002; Sanders 2006) at specific wavelengths in order to define physically-motivated regions. More recently, Sánchez et al. (2016) proposed an alternative method, also aiming to obtain a tessellation with a target S/N, that imposes ‘continuity’ in the surface brightness (i.e. a maximum contrast within any region) in order to better adapt to the morphology of the data.

This work takes another step in this direction, presenting an alternative approach that is not aimed to obtain a specific S/N but to identify spatial regions where the signal is statistically consistent with being constant within the provided errors: if two regions carry the same information, it will always be advantageous to merge them in order to (further) increase the signal-to-noise ratio; if they are ‘different’ (do not carry the same information), they should be kept separate in order not to introduce artificial biases. Such prescription preserves the information contained in the input dataset, and it imposes no condition on the shape of the tessellation. It is extremely general, and it can be applied to any kind of spatially-resolved data.

We describe the mathematical basis of the method in Section 2, while the details of the algorithm and its implementation are discussed in Section 3. Section 4 proposes a set of benchmark problems, and the analysis of the results is discussed in Section 5. Our main conclusions are summarized in Section 6.

2 MATHEMATICAL FORMULATION OF THE PROBLEM

BATMAN (Bayesian Technique for Multi-image Analysis) is a new binning scheme designed to perform an adaptive segmentation on a three-dimensional dataset that serves as

inceptive step for its further analysis. We will refer to these sets as ‘datacubes’ or ‘multi-images’ along the manuscript, as they consist of an arbitrary number (n_λ ‘layers’ or ‘wavelengths’) of spatially-arranged two-dimensional data (individual ‘images’ or ‘maps’). We will focus on IFS data throughout the paper, and we will use the term ‘spaxel’ to denote the individual and minimum spatial element of a dataset. However, the procedure can be applied to any set of images, or to any kind of spatially-distributed data. In particular, the algorithm can also be applied to a single image (i.e. single-‘layer’, 2D dataset) in what we will refer to as ‘monochromatic’ binning mode.

More precisely, BATMAN works on an *input dataset* $D = \{X, E\}$ consisting of two $n_{\text{row}} \times n_{\text{col}}$ multi-images X and E , with $N = n_{\text{row}} \times n_{\text{col}}$ spaxels each, whose spatial positions we will tag with the indices $\{i, j\}$. Every spaxel has a set of n_λ ‘measurements’ $x_{ij\lambda}$ and ‘errors’ $e_{ij\lambda}$

$$\begin{aligned} X &= \{x_{ij\lambda}\} \\ E &= \{e_{ij\lambda}\} \end{aligned} \quad (1)$$

associated to its spatial location, where $i = [1, n_{\text{row}}]$, $j = [1, n_{\text{col}}]$, and $\lambda = [1, n_\lambda]$. Different ‘layers’ may correspond to broad-band observations of the same object, different wavelengths in an IFS datacube, or even maps with completely different information (e.g. mass surface density, velocity, velocity dispersion, age, metallicity, equivalent width, etc).

The goal of the algorithm is to divide the input dataset into spatially-connected ‘regions’ that carry the same information, i.e. where all the measurements $x_{ij\lambda}$ are statistically compatible with being random samples of the same underlying signal $S_{ij\lambda}$ given the Gaussian errors $e_{ij\lambda}$. In the current implementation, BATMAN makes use of Bayesian statistics to connect adjacent spaxels into ‘regions’ (and merge adjacent regions into larger ones) as long as it is more likely that the true values of the measured quantities (the ‘signal’) are constant within the merged region ($S_{ij\lambda} = \theta_{r\lambda} \forall \{i, j\}$ in region r) than a collection of several independent domains with different values. More elaborate (e.g. quadratic) models of the underlying signal may be considered in future versions of the code.

Thus, the *output dataset* is an optimized ‘segmentation model’ $M = \{R, \Theta\}$ consisting of n_{reg} connected regions (with every spaxel $\{i, j\}$ assigned to the region r_{ij}) where the signal at the different bands is assumed to be well described by the constant values $\theta_{r\lambda}$:

$$\begin{aligned} R &= \{r_{ij}\} \\ \Theta &= \{\theta_{r\lambda}\} \end{aligned} \quad (2)$$

with $r = [1, n_{\text{reg}}]$. R refers thus to the tessellation of the image, fully specified by a $n_{\text{row}} \times n_{\text{col}}$ matrix containing the labels r_{ij} assigned to every spaxel, and Θ corresponds to the set of $n_{\text{reg}} \times n_\lambda$ parameters $\theta_{r\lambda}$ used to describe the signal.

According to the frequentist approach, the best segmentation model would be given by the values of R and Θ that maximise the *likelihood*, i.e. the probability of obtaining the measurements X , given the errors E and the n_{reg} regions specified by R and Θ . The model proposed for this work assumes that the observed values $x_{ij\lambda}$ are statistically-independent random variables, with Gaussian probability distributions whose dispersions are given by the errors $e_{ij\lambda}$ while their means $\theta_{r_{ij}\lambda}$ are to be determined. With these

	Symbol	Meaning	Equation
DIMENSIONS	i, j λ	spatial dimensions (bi-dimensional, $n_{\text{row}} \times n_{\text{col}}$ maps) spectral dimension (n_λ ‘layers’ or ‘wavelengths’)	
INPUT	$D = \{X, E\}$ $X = \{x_{ij\lambda}\}$ $E = \{e_{ij\lambda}\}$	input data set (measurements and errors) given to BATMAN measurements at every 3D position errors associated to every 3D position	(1)
MODEL	$M = \{R, \Theta\}$ $R = \{r_{ij}\}$ $\Theta = \{\theta_{r\lambda}\}$ $\mathcal{L}(X E, R, \Theta)$	assumed model (tessellation and signals) to describe the data tessellation (labels associated to every spatial position) signal associated to every region r and wavelength λ likelihood of the measurements, given the errors and the model	(2) (3)
PRIORS	$\mathcal{P}(R, \Theta) = \mathcal{P}(R) \mathcal{P}(\Theta R)$ $\mathcal{P}(\Theta R)$ $\mathcal{P}(R)$	prior probability distribution of the model parameters prior probability of the signals Θ for a given tessellation R prior probability of the tessellations R	(7) (15)
POSTERIORES	$p(R, \Theta D)$ $p(\Theta R, D)$ $p(R D)$ $\mu_{r\lambda}$ $\sigma_{r\lambda}$	posterior probability of the model parameters, given the data posterior probability of the signals Θ , given R and D posterior probability of the tessellations R , given D expected value of the signal according to $p(\Theta R, D)$ expected variance (‘formal error’) of the expected signal	(4) (8), (10) (16) (11) (12)
EVIDENCE	\mathcal{E} $\mathcal{E}(R)$	overall Bayesian evidence for our assumed description evidence for the tessellation R	(5), (17) (9), (13)
NUMERICAL IMPLEMENTATION	K R_n $R_{c,n-1}$	combined prior parameter currently-accepted tessellation with n regions candidate tessellations with $n - 1$ regions	(20) (19)
TEST CASES AND ANALYSIS	$S_{ij\lambda}$ $e_{ij\lambda}$ $e_{\text{ccd}}, e_{\text{uniform}}$ $\Delta_{ij\lambda}$ $\hat{\Delta}_{ij\lambda}$	signal associated to every 3D position (Eq. 22) noise associated to every 3D position (Eq. 22) different error prescriptions, see Section 4.1 residual shown in Figure 3 (Eq. 26) residual shown in Figure 4 (Eq. 26)	

Table 1. List of symbols used in the article.

considerations, one can express the likelihood as

$$\begin{aligned}
 \mathcal{L}(X|E, R, \Theta) &= \prod_{i=1}^{n_{\text{row}}} \prod_{j=1}^{n_{\text{col}}} \prod_{\lambda=1}^{n_\lambda} e^{-\frac{(x_{ij\lambda} - \theta_{r_{ij}\lambda})^2}{2e_{ij\lambda}^2}} \\
 &= \prod_{r=1}^{n_{\text{reg}}} \prod_{i,j \in r} \prod_{\lambda=1}^{n_\lambda} e^{-\frac{(x_{ij\lambda} - \theta_{r\lambda})^2}{2e_{ij\lambda}^2}}
 \end{aligned} \quad (3)$$

where the second product in the last term simply loops over the spaxels in a given region.

Within the Bayesian framework BATMAN is based upon, the approach is slightly different, and it comprises several additional ingredients. One must declare not only a likelihood for the data given the model, but also a *prior* probability distribution $\mathcal{P}(R, \Theta)$ of the model parameters. The construction of this prior is based on our previous knowledge about the problem, and it will certainly affect the inferences we obtain from our data. As an example, it could favour certain tessellations (roundish bin shapes, such as in Cappellari & Copin 2003; Sánchez et al. 2012b, 2016) or avoid/penalize unphysical values (e.g. negative values of the signal).

The Bayesian approach combines the prior with the likelihood of observing the measurements X (making use of Bayes’ theorem) to compute the *posterior* probability distribution of the model parameters

$$p(R, \Theta|D) = \frac{\mathcal{P}(R, \Theta) \mathcal{L}(X|E, R, \Theta)}{\mathcal{E}} \quad (4)$$

given the data. The overall Bayesian *evidence* for our as-

sumed description

$$\mathcal{E} \equiv p(X|E) = \sum_R \int \mathcal{P}(R, \Theta) \mathcal{L}(X|E, R, \Theta) d\Theta \quad (5)$$

is the sum over all possible tessellations R and signal values Θ , and it can be considered as the overall probability of observing the measurements X , given the errors E , if our model provided an accurate description of the data (in frequentist terms, the likelihood of our description). The expected values of the model parameters and their uncertainties can be computed from the posterior probability distribution (4).

In BATMAN, the problem of image segmentation is split in two parts:

- (i) the estimation of the signal distribution Θ given a fixed segmentation R
- (ii) the further selection of the most probable tessellation

The exact procedures followed to solve every one of them are discussed in the following sections.

2.1 Parameter estimation

Given a fixed tessellation R , BATMAN will evaluate the posterior probability distribution $p(\Theta|R, D)$ describing the signal $\theta_{r\lambda}$ in every region r at every wavelength λ .

Following a Bayesian procedure, we first assume that the measurements X are described by the Gaussian likelihood function (see Eq. 3). Then, as we do not expect any specific distribution for the values of $\theta_{r\lambda} \in \Theta$, we have adopted a uniform (sometimes referred to as ‘noninformative’ or ‘objective’) prior. The uniform probability distribu-

tion assigns equal probability to all values of the parameter, reflecting our initial ignorance. In order to obtain a *proper* prior, for which the total probability is normalized to unity (in contrast to an *improper* prior, where the integral diverges to infinity), the possible values of the signal $\theta_{r\lambda}$ are restricted to a finite range of values:

$$\mathcal{P}(\theta_{r\lambda}|R) d\theta_{r\lambda} = \frac{1}{H_\lambda - L_\lambda} d\theta_{r\lambda} \quad (6)$$

$$L_\lambda < \theta_{r\lambda} < H_\lambda$$

where upper and lower limits H_λ and L_λ correspond to the highest and lowest values measured at that particular wavelength, anywhere within the image (i.e. the maximum and minimum $x_{ij\lambda}$ over i and j for a given λ).

From a practical point of view, this prescription automatically selects a reasonable range for the signal (irrespective of e.g. measurement units in the different layers of the multi-image), but, from a strict Bayesian perspective, one may object that we have actually looked at the data in order to set our prior. In order to generalise expression (6), we introduce n_λ additional parameters $0 < k_\lambda \leq 1$ such that

$$\begin{aligned} \mathcal{P}(\Theta|R) d\Theta &= \prod_{\lambda=1}^{n_\lambda} \prod_{r=1}^{n_{\text{reg}}} \mathcal{P}(\theta_{r\lambda}|R) d\theta_{r\lambda} \\ &= \prod_{\lambda=1}^{n_\lambda} \left(\frac{k_\lambda}{H_\lambda - L_\lambda} \right)^{n_{\text{reg}}} d\theta_{r\lambda} \quad (7) \\ L_\lambda - \Delta_\lambda &< \theta_{r\lambda} < H_\lambda + \Delta_\lambda \end{aligned}$$

where $\Delta_\lambda = \frac{1-k_\lambda}{2k_\lambda}(H_\lambda - L_\lambda)$. The choice $k_\lambda = 1$ results in equation (6), whereas $k_\lambda = 0$ would yield the improper prior $-\infty < \theta_{r\lambda} < \infty$. Thus, the value of k_λ provides a qualitative indication of the weight given to the information contained in H_λ and L_λ .

Using Bayes' theorem, the posterior probability distribution for the signal is given by

$$p(\Theta|R, D) d\Theta = \frac{\mathcal{P}(\Theta|R) \mathcal{L}(X|E, R, \Theta)}{\mathcal{E}(R)} d\Theta \quad (8)$$

where $\mathcal{L}(X|E, R, \Theta)$ is the likelihood defined in equation (3), and the evidence for the tessellation R can be expressed as

$$\mathcal{E}(R) = p(X|E, R) = \int \mathcal{P}(\Theta|R) \mathcal{L}(X|E, R, \Theta) d\Theta \quad (9)$$

Since we have adopted a uniform prior, and our model likelihood is a product of $n_{\text{row}} \times n_{\text{col}} \times n_\lambda$ Gaussians, it is easy to verify¹ that the posterior probabilities

$$p(\theta_{r\lambda}|R, D) \propto e^{-\frac{(\theta_{r\lambda} - \mu_{r\lambda})^2}{2\sigma_{r\lambda}^2}} \quad (10)$$

are independent Gaussians with mean

$$\mu_{r\lambda} = \sigma_{r\lambda}^2 \sum_{i,j \in r} \frac{x_{ij\lambda}}{e_{ij\lambda}^2} \quad (11)$$

and dispersion

$$\frac{1}{\sigma_{r\lambda}^2} = \sum_{i,j \in r} \frac{1}{e_{ij\lambda}^2} \quad (12)$$

¹ Using that $\sum_i \frac{(x-\mu_i)^2}{2\sigma_i^2} = \frac{(x-\mu)^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \sum_i \frac{\mu_i^2}{2\sigma_i^2}$ with $\frac{1}{\sigma^2} \equiv \sum_i \frac{1}{\sigma_i^2}$ and $\mu \equiv \sigma^2 \sum_i \frac{\mu_i}{\sigma_i^2}$.

for any value of the prior parameters k_λ .

In other words, the expected value $\mu_{r\lambda}$ of the signal within any given region r reported by BATMAN (and the associated ‘formal errors’ $\sigma_{r\lambda}$) are obtained from an *inverse-variance weighted average* over the region.

Finally, the evidence for the tessellation R is given by

$$\begin{aligned} \mathcal{E}(R) &= \int \prod_{r=1}^{n_{\text{reg}}} \prod_{\lambda=1}^{n_\lambda} \mathcal{P}(\theta_{r\lambda}|R) \prod_{i,j \in r} \frac{e^{-\frac{(x_{ij\lambda} - \theta_{r\lambda})^2}{2e_{ij\lambda}^2}}}{\sqrt{2\pi e_{ij\lambda}^2}} d\theta_{r\lambda} \\ &= \prod_{r=1}^{n_{\text{reg}}} \prod_{\lambda=1}^{n_\lambda} \frac{k_\lambda}{H_\lambda - L_\lambda} \frac{\int e^{-\sum_{i,j \in r} \frac{(x_{ij\lambda} - \theta_{r\lambda})^2}{2e_{ij\lambda}^2}} d\theta_{r\lambda}}{\prod_{i,j \in r} \sqrt{2\pi e_{ij\lambda}^2}} \quad (13) \\ &\equiv \prod_{r=1}^{n_{\text{reg}}} \mathcal{E}_r \end{aligned}$$

with

$$\begin{aligned} \mathcal{E}_r &\equiv \prod_{\lambda=1}^{n_\lambda} \frac{k_\lambda \int e^{-\frac{(\theta_{r\lambda} - \mu_{r\lambda})^2}{2\sigma_{r\lambda}^2} + \frac{\mu_{r\lambda}^2}{2\sigma_{r\lambda}^2} - \sum_{i,j \in r} \frac{x_{ij\lambda}^2}{2e_{ij\lambda}^2}} d\theta_{r\lambda}}{(H_\lambda - L_\lambda) \prod_{i,j \in r} \sqrt{2\pi e_{ij\lambda}^2}} \quad (14) \\ &= \prod_{\lambda=1}^{n_\lambda} \frac{k_\lambda e^{\frac{\mu_{r\lambda}^2}{2\sigma_{r\lambda}^2} - \sum_{i,j \in r} \frac{x_{ij\lambda}^2}{2e_{ij\lambda}^2}} \sqrt{2\pi\sigma_{r\lambda}^2}}{(H_\lambda - L_\lambda) \prod_{i,j \in r} \sqrt{2\pi e_{ij\lambda}^2}} \end{aligned}$$

2.2 Model selection

Once we obtain the evidence $\mathcal{E}(R)$ for any tessellation R and the posterior probability distribution of the recovered signal Θ associated to R , we face the question of selecting the ‘optimal’ segmentation that is most likely to describe our dataset. This is the second time we apply a Bayesian approach to our dataset, but two main differences exist with respect to the parameter estimation problem described in the previous section:

- (i) the set of all possible tessellations is discrete, whereas the values of the signal are continuous variables
- (ii) we will require BATMAN to select one *and only one* ‘optimal’ tessellation rather than a probabilistic description (which is, as we will discuss later, not ‘strictly Bayesian’).

Except for these two differences, the Bayesian approach to model selection is very similar to parameter estimation: first, we must specify our priors $\mathcal{P}(R)$ and then combine them with the appropriate likelihood $p(X|E, R)$ in order to obtain the posterior probability distribution $p(R|D)$ from Bayes' theorem.

We have no preliminary information relative to the preferred tessellations, and BATMAN will not impose any geometrical constraint (e.g. roundness) other than ensuring that all the regions defined by the matrix r_{ij} are physically connected. It is however foreseen that some practical applications of the algorithm may have preference for a larger or smaller number of regions n_{reg} , and therefore we have chosen a prior of the form

$$\mathcal{P}(R) = k_R^{n_{\text{reg}}} \quad (15)$$

where k_R is to be defined by the user. The value $k_R = 1$ corresponds to the improper uniform prior, that assigns equal

probability to all valid segmentations regardless of the number of regions they contain; $k_R < 1$ would favour that the multi-image is divided into a small number of (large) regions, whereas $k_R > 1$ would favour that individual spaxels are kept independent or cluster into a large number of small regions.

The likelihood of a tessellation R (i.e. the probability of measuring X , considering all possible values of Θ) is simply the evidence $\mathcal{E}(R)$ defined in equation (9), and thus the posterior probability distribution of R is given by

$$p(R|D) = \frac{\mathcal{P}(R) \mathcal{E}(R)}{\mathcal{E}} \quad (16)$$

where \mathcal{E} denotes the overall evidence (5). Due to the discrete nature of R ,

$$\mathcal{E} = \sum_R \mathcal{P}(R) \mathcal{E}(R) \quad (17)$$

where the R runs over all possible tessellations.

The current version of BATMAN outputs the matrix r_{ij} (labels corresponding to the tessellation) that maximises the posterior probability $p(R|D)$, along with the corresponding values of $\mu_{r\lambda}$ and $\sigma_{r\lambda}$. Let us argue at this point that, to some extent, the very concept of ‘model selection’ is intrinsically not Bayesian. If, for example, $p(R_+|D) = 0.5 + \delta$ and $p(R_-|D) = 0.5 - \delta$, one can only choose R_+ , no matter how small δ may be, if forced to take a decision. We do think that a strictly Bayesian algorithm would never select a unique segmentation, but calculate the posterior probability of all possible tessellations and weight them in order to provide a fully probabilistic description of the underlying signal, spaxel by spaxel. However, such approach would yield a ‘smoothing’ technique rather than a ‘segmentation’ tool, and it will not be discussed further in the present work.

3 NUMERICAL IMPLEMENTATION

BATMAN is based on the philosophy that the purpose of a segmentation algorithm is to group together the ‘spaxels’ of a ‘multi-image’ that carry the same information (measurements $x_{ij\lambda}$ compatible with identical signal $S_{ij\lambda}$ within the errors $e_{ij\lambda}$) into larger ‘regions’. According to the arguments presented in Section 2, this goal can be mathematically formulated as the maximisation of the Bayesian posterior probability of the tessellation R , fully described by the $n_{\text{row}} \times n_{\text{col}}$ matrix r_{ij} specifying the label of the region that each spaxel (i, j) belongs to. Once R is fixed, we have shown that the posterior probability distribution of the ‘signal’ $\theta_{r\lambda}$ within each region r at ‘wavelength’ (or ‘layer’) λ is Gaussian, with mean $\mu_{r\lambda}$ and standard deviation $\sigma_{r\lambda}$ given by an inverse-variance weighted average over the region, i.e. equations (11) and (12), respectively.

In practice, the number of possible tessellations is so large that evaluating all their evidence is completely infeasible. Therefore, BATMAN follows a greedy iterative procedure, merging adjacent regions until no further increase in the posterior probability $p(R|D)$ is possible:

(i) A unique label $l_{ij} = [1, N]$ is assigned to each spaxel. The initial segmentation R_N considers that every spaxel is an independent region, i.e. $n_{\text{reg}} = N$, $r_{ij} = l_{ij}$, $\mu_{r\lambda} = x_{ij\lambda}$,

and $\sigma_{r\lambda} = e_{ij\lambda}$. From now on, we will drop the subscript and denote the number of regions simply as n .

(ii) On every iteration, starting from $n = N$, BATMAN compares the posterior probability of R_n with all possible candidate tessellations $R_{c,n-1}$ that can be obtained by merging any two adjacent² regions in R_n . More precisely, it evaluates the ratios $\frac{p(R_{c,n-1}|D)}{p(R_n|D)}$ for all these candidate tessellations with $n - 1$ regions each, where the subscript c runs over all the candidates considered.

(iii) The algorithm selects the optimal tessellation for the next iteration, R_{n-1} , as the candidate displaying the highest probability (i.e. the largest ratio with respect to R_n). If none of the $R_{c,n-1}$ is found to be more likely than R_n , no further iterations will be performed. The greedy procedure is illustrated in Appendix A, and the exit condition is discussed in Appendix B.

(iv) On exit, BATMAN outputs the matrices containing the labels r_{ij} of the different regions, the mean posterior signal maps $\mu_{r\lambda}$, and their standard deviation $\sigma_{r\lambda}$ for the final ‘optimized’ tessellation R_n .

A critical part of the algorithm is thus the evaluation of the probability ratios

$$\begin{aligned} \frac{p(R_{c,n-1}|D)}{p(R_n|D)} &= \frac{\mathcal{P}(R_{c,n-1}) \mathcal{E}(R_{c,n-1})}{\mathcal{P}(R_n) \mathcal{E}(R_n)} \\ &= \frac{\mathcal{P}(R_{c,n-1}) \prod_{r' \in R_{c,n-1}} \mathcal{E}_{r'}}{\mathcal{P}(R_n) \prod_{r \in R_n} \mathcal{E}_r} \end{aligned} \quad (18)$$

where r and r' refer to the regions defined in R_n and $R_{c,n-1}$, respectively. Due to our iterative procedure, these two models only differ in that two regions of R_n (let us name them A and B) that have been merged into a single region $A \cup B$ in $R_{c,n-1}$. Substituting expression (14) for the contribution of each region to the evidence and equation (15) for the prior probability of each tessellation, one arrives to

$$\begin{aligned} \frac{p(R_{c,n-1}|D)}{p(R_n|D)} &= \frac{k_R^{n-1}}{k_R^n} \frac{\mathcal{E}_{A \cup B}}{\mathcal{E}_A \mathcal{E}_B} \\ &= k_R^{-1} \prod_{\lambda=1}^{n_\lambda} \frac{(H_\lambda - L_\lambda) \sigma_{A \cup B, \lambda} e^{\frac{\mu_{A \cup B, \lambda}^2}{2\sigma_{A \cup B, \lambda}^2}}}{k_\lambda \sqrt{2\pi} \sigma_{A\lambda} \sigma_{B\lambda} e^{\frac{\mu_{A\lambda}^2}{2\sigma_{A\lambda}^2} + \frac{\mu_{B\lambda}^2}{2\sigma_{B\lambda}^2}}} \\ &\equiv K^{-1} \prod_{\lambda=1}^{n_\lambda} (H_\lambda - L_\lambda) \frac{e^{-\frac{(\mu_{A\lambda} - \mu_{B\lambda})^2}{2(\sigma_{A\lambda}^2 + \sigma_{B\lambda}^2)}}}{\sqrt{2\pi(\sigma_{A\lambda}^2 + \sigma_{B\lambda}^2)}} \end{aligned} \quad (19)$$

where the prior parameters k_R and k_λ can be neatly separated from the term comparing the measurements in regions A and B in terms of their dispersions. In fact, it is evident from expression (19) that the combined effect of k_R and k_λ can be grouped into a single number

$$K \equiv k_R \prod_{\lambda=1}^{n_\lambda} k_\lambda \quad (20)$$

that encapsulates all our choices about the prior probability distributions for r_{ij} and $\theta_{r\lambda}$. Since all other quantities

² We consider contiguous spaxels in the horizontal or vertical directions, but not along the diagonals.

(including H_λ and L_λ , which denote the maximum and minimum values of $x_{ij\lambda}$, respectively) are driven by the data, K is the only free parameter of our algorithm.

4 TEST CASES

In order to illustrate the capabilities of the algorithm, assess its performance, and provide general guidelines to understand its operation, we have considered a set of four different input datasets as benchmark problems. Two of them are based on a synthetic multi-image (Section 4.1) that poses several potentially challenging situations (e.g. different shapes, with and without sharp boundaries, different noise statistics and signal-to-noise ratios, overlap between independent structures, etc.), and the other two are based on integral-field spectroscopic observations of the local galaxy NGC 2906 (Section 4.2).

These four datasets consist of several ‘layers’ or ‘wavelengths’ (i.e. they are all ‘multi-images’ according to our terminology). For every one of them we have run BATMAN in two different ways: one considering all the layers simultaneously and obtaining a single common tessellation (we will refer to this approach as ‘multi- λ ’) and another in which we consider every ‘layer’ individually and obtain an independent segmentation for every one of them (‘monochromatic’ approach). This yields a total of eight test cases, as summarized in Table 3.

For all of them, we carry out two independent runs adopting $K = 1$ and $K = 10^{-6}$. The former (obtained e.g. by setting both k_R and every k_λ to unity) corresponds to a binning criterion that tends to keep small regions separated; reducing the value of K (e.g. by decreasing either k_R and/or the product of the different k_λ) favours the coaddition of spaxels and leads to the definition of larger regions.

4.1 Synthetic data

Our synthetic multi-images consist of 3 layers (arbitrarily labelled R, G, and B) that display different objects over a null background where the signal value is set to $S_0 = (0, 0, 0)$. The R layer shows a quarter of a circle in the bottom-left corner with signal $S_{\text{circle}} = (1, 0, 0)$, the G layer displays a triangle with $S_{\text{triangle}} = (0, 7, 0)$ located the centre of the image, slightly overlapping with the circle, and the B image shows an ellipse centred in the top-right corner (again, slightly overlapping with the triangle) that displays a continuous, radially-decreasing gradient $S_{\text{ellipse}} = (0, 0, [17-1])$.

This three-layer datacube

$$S = (S_R, S_G, S_B) = S_0 + S_{\text{circle}} + S_{\text{triangle}} + S_{\text{ellipse}} \quad (21)$$

represents the *signal* of our synthetic problem. From these data, we generate two different *input* multi-images

$$x_{ij\lambda} = S_{ij\lambda} + e_{ij\lambda} \quad (22)$$

by adding random Gaussian noise n with different statistics. The ‘uniform-noise’ datacube follows a Gaussian probability distribution with zero mean and dispersion $e_{\text{uniform}} = (1, 1, 1)$ for all layers of every spaxel. We also generate a ‘CCD-like’ datacube (mimicking the characteristics of charge-coupled devices) as a result of adding a white-noise

	S	e_{ccd}	$(S/N)_{\text{ccd}}$
Background	0	0.5	0
R: Circle	1	1	1
G: Triangle	7	2	3.5
B: Ellipse	[1 – 17]	[1 – 3]	[1 – 5.67]

Table 2. Values of the signal S of each individual component of the synthetic test multi-image (21) in the relevant layer (R, G, or B), followed by the corresponding ‘CCD-like’ noise e_{ccd} given by expression (23) and the signal-to-noise ratio $(S/N)_{\text{ccd}} = S/e_{\text{ccd}}$. For the ‘uniform-noise’ datacube, $e_{\text{uniform}} = (1, 1, 1)$ implies $(S/N)_{\text{uniform}} = S/e_{\text{uniform}} = S$.

component with dispersion $\sigma_W^2 = (0.5, 0.5, 0.5)$ and a Poisson noise component where the dispersion σ_P^2 grows linearly with the signal $S_{ij\lambda}$ at any given band and location, according to the expression

$$e_{\text{ccd};ij\lambda}^2 = \sigma_W^2 + \sigma_P^2 = \frac{1 + S_{ij\lambda}}{2} \quad (23)$$

The numeric values of the intensities and noise levels in our synthetic test problems have been chosen so that the signal-to-noise ratio

$$(S/N)_{ij\lambda} = S_{ij\lambda}/e_{ij\lambda} \quad (24)$$

covers a range from one to 17 in the ‘uniform’ case and up to approximately 6 for the ‘CCD-like’ noise. The precise values for the errors and signal-to-noise ratios within each region are summarized in Table 2.

4.2 Astronomical observations

In order to illustrate the use of the algorithm in a typical science case, we consider two different segmentation problems related to the measurement of the intensity of the Balmer emission lines in the local spiral galaxy NGC 2906. Observations have been retrieved from the public database of the Calar Alto Legacy Integral-Field Spectroscopic Area (CALIFA) survey (Sánchez et al. 2012a). More precisely, they correspond to the COMBO datacubes delivered in Data Release 2 (see García-Benito et al. 2015, for a detailed description), with a uniform wavelength coverage from 3650 to 7200 Å, sampled in constant steps of 2 Å.

We will consider two completely different approaches to address the problem. In the first one, we measure the intensity of the $H\alpha(6563\text{Å})$, $H\beta(4861\text{Å})$, $H\gamma(4340\text{Å})$, and $H\delta(4101\text{Å})$ emission lines spaxel by spaxel, and then take the resulting intensity maps, with their corresponding errors, as a four-layer input multi-image (i.e. first measure and then bin with BATMAN). As an alternative approach, we run BATMAN directly on the raw IFS data, selecting four wavelength intervals of ± 15 Å around each line (see Table 3), and then measure their intensity from the integrated spectrum within each region (i.e. first bin the datacube directly with BATMAN and then measure).

In both cases, we follow exactly the same procedure to measure the intensity of the Balmer lines in a given spectrum. In order to account for stellar absorption lines, we use STARLIGHT (Cid Fernandes et al. 2005) to model the spectral energy distribution of the underlying stellar population, fitting the observed spectrum with a linear combination of simple stellar populations spanning different ages and metallicities. We use the spectra provided by Vazdekis et al.

	TEST CASE	INPUT: ($n_{\text{row}} \times n_{\text{col}} \times n_{\lambda}$)	n_{output}	FIGURES	PANELS
#1	synthetic + uniform noise 'monochromatic'	R: ($73 \times 78 \times 1$) G: ($73 \times 78 \times 1$) B: ($73 \times 78 \times 1$)	3	1, 3, 4	top-right
#2	synthetic + uniform noise 'multi- λ '	($73 \times 78 \times 3$)	1	1, 3, 4	top-middle
#3	synthetic + ccd-like noise 'monochromatic'	R: ($73 \times 78 \times 1$) G: ($73 \times 78 \times 1$) B: ($73 \times 78 \times 1$)	3	1, 3, 4	bottom-right
#4	synthetic + ccd-like noise 'multi- λ '	($73 \times 78 \times 3$)	1	1, 3, 4	bottom-middle
#5	NGC 2906: SHIFU maps 'monochromatic'	H α : ($71 \times 78 \times 1$) H β : ($71 \times 78 \times 1$) H γ : ($71 \times 78 \times 1$) H δ : ($71 \times 78 \times 1$)	4	2, 5	top-right
#6	NGC 2906: SHIFU maps 'multi- λ '	($71 \times 78 \times 4$)	1	2, 5	top-middle
#7	NGC 2906: CALIFA data ' $\Delta\lambda = \pm 15\text{\AA}$ '	H α : ($71 \times 78 \times 14$) H β : ($71 \times 78 \times 15$) H γ : ($71 \times 78 \times 15$) H δ : ($71 \times 78 \times 15$)	4	2, 5	bottom-right
#8	NGC 2906: CALIFA data 'full set'	($71 \times 78 \times 59$)	1	2, 5	bottom-middle

Table 3. Summary of the test cases presented in Section 4 and discussed in Section 5. The labelling and nomenclature used to refer each case are quoted on the first two columns, respectively. The dimensionality of the input data is listed on the third column, whereas the fourth column provides the number of output tessellations (i.e. the number of times BATMAN is run), and the last two columns indicate the location of the corresponding results in the article figures. Every test case has been run with two different values for the prior parameter, $K = 1$ and $K = 10^{-6}$.

(2010) for populations older than 64 Myr and González Delgado et al. (2005) models for younger ages. Dust effects are modelled as a foreground screen, assuming a Cardelli et al. (1989) reddening law with $R_V = 3.1$. Then, we subtract the estimated stellar continuum from the observed spectrum and obtain the line flux with the Sherpa IFU line fitting software (SHIFU; García-Benito, in preparation), based on CIAO's SHERPA package (Freeman et al. 2001; Doe et al. 2007). Small deviations with respect to the stellar continuum are taken into account by a first order polynomial, and independent Gaussians have been fitted for the emission lines.

The intensity distribution of the Balmer emission lines in NGC 2906 is fairly clumpy (see Figure 2), with a relatively large number of individual compact HII regions arranged in a ring-like structure. This configuration is clearly visible in the H α maps, where the signal is stronger, but it is much more difficult to identify in the weakest lines, especially H δ . In addition, there is a weak diffuse component of much lower surface brightness, arising from a combination of intrinsic emission from diffuse ionized gas (Haffner et al. 2009) and light from the compact HII regions that is scattered towards the observer by dust particles in the interstellar medium (see e.g. Ascasibar et al. 2016).

Although the actual solution is obviously not available in this case, all our emission lines arise from the same element, and therefore they should all roughly trace the same spatial distribution (mostly determined by the electron number density, with a secondary dependence on the local electron temperature). However, the intensity of the lines decreases with frequency, and therefore they probe very different signal-to-noise ratios, from values of the order unity,

especially in the outer parts, to well above one hundred in the regions of brightest H α emission.

4.3 Analysis procedure

As mentioned above, we have approached our four input datasets (both synthetic and astronomical) in two different ways. On the one hand, we make use of the default binning mode ('multi- λ ') where all the information is considered at the same time (all 'layers' binned simultaneously) and one optimal tessellation is computed for the whole dataset. On the other hand, an alternative 'monochromatic' procedure is also considered, where BATMAN is run individually on every layer of the multi-image dataset (hence, several times) and an independent tessellation is derived for every one of them. A summary of all the resulting test cases is provided in Table 3. Every test has been repeated twice, adopting the values $K = 1$ and $K = 10^{-6}$ for the combined prior parameter.

In the synthetic multi-images, BATMAN has been applied to the three layers (R, G, and B) separately ('monochromatic', cases #1 and #3) and simultaneously ('multi- λ ', cases #2 and #4). The two test cases based on SHIFU measurements of NGC 2906 (cases #5 and #6) follow an approach completely analogous to the synthetic multi-images. The only difference is that now the data consist of four different (but physically not independent) 'layers', one for each Balmer line. In case #5, BATMAN is run in 'monochromatic' mode, producing 4 different segmentations, while in case #6 it is run only once, yielding a single 'multi- λ ' tessellation. For the remaining observational test cases, #7 and #8, we have applied BATMAN directly on the IFS data from the CALIFA survey (spectra and error). We have

run our algorithm separately on every 30 Å-wide slice around each line, thus obtaining 4 independent tessellations (test case #7), as well as to the full 59-layer multi-image (the four 30 Å-wide slices together, case #8), where a single tessellation is returned. Although these tests are scientifically identical to test cases #5 and #6 (4 balmer lines binned separately and simultaneously), they are technically different: BATMAN is applied in ‘multi-image’ mode every time, many more ‘wavelengths’ are being considered, and there is no pre-processing of the data (lines are measured with SHIFU after the binning).

As a final remark, let us note that there is another important difference between test cases #7 and #8 (direct application of the algorithm on the IFS datacube) and the rest. In test cases #1 – 6, the input data are maps of the measured quantities. BATMAN provides not only an optimal tessellation, but also the expected value $\mu_{r\lambda}$ and the dispersion $\sigma_{r\lambda}$ of the signal within every region, based on the posterior probability distribution. On the contrary, test cases #7 and #8 bin directly a selected section of the spectral energy distributions. Although BATMAN returns again the three products (r_{ij} , $\mu_{r\lambda}$ and $\sigma_{r\lambda}$) only the tessellation is further used by SHIFU in order to derive the flux of each line and the corresponding error, taking into account noise covariance in CALIFA data (see e.g. Husemann et al. 2013; García-Benito et al. 2015) as part of its pipeline. Therefore, the estimated errors may be considerably larger (by about a factor of 3, depending on the number of spaxels within each region) than those implied by equation (12).

5 RESULTS

Based on the test cases described in the previous section, we now address the quality of the output tessellation in terms of its ability to adapt to the underlying signal, the dependence of the results on the adopted priors, the accuracy of the formal errors estimated by the algorithm, and the improvement with respect to the original input data.

5.1 Large-scale morphology

Figures 1 and 2 show the results of applying BATMAN to our synthetic and astrophysical test cases, respectively, adopting the extreme values $K = 1$ and $K = 10^{-6}$ for the prior parameter, and considering both the full input data set as well as individual layers in order to compute the Bayesian evidence (see Table 3). From visual inspection, one can readily verify that, in almost all cases, the tessellations returned by the algorithm adapt extremely well to the structures that were present in the input data.

In the synthetic datacubes (Figure 1), BATMAN successfully recovers the different objects regardless of their topology. The output tessellations present a variety of sizes and shapes that are driven by the input signal, and they show no preference for any given direction. Comparing the results obtained for the ‘uniform’ and ‘CCD-like’ noise schemes, we conclude that the ability of the algorithm to trace the underlying structure is not very sensitive to the statistical characteristics of the noise. The minor differences that exist in the recovered tessellations can be explained in terms of the dissimilar signal-to-noise ratios.

When applied to the real data (Figure 2), our algorithm manages to identify most of the clumps associated with physical HII regions, albeit with significant difficulty in the case of H δ . In particular, among the H δ recovered maps, the poorest results are obtained for case #5, where only the H δ intensity and its error map, as computed by SHIFU, are used as input dataset. Interestingly, the situation improves when BATMAN is run directly on the CALIFA COMBO dataset. Even when only a narrow spectral region, ‘ $\Delta = \pm 15$ Å’, around H δ is provided as input (case #7), the ring-like structure where most of the emission comes from is prominently visible in the recovered signal.

By construction, BATMAN gathers together spaxels that contain the same information (compatible signal within the errors). Hence, it is reassuring, but not surprising, that it successfully recovers the circle and the triangle in the R and G layers of the synthetic multi-images, as these regions do indeed feature exactly the same signal value. However, this is not the case for the gradient in the B-band ellipse or the observations of NGC 2906, where the signal varies more or less smoothly across any layer. Our segmentation model, described in Section 2, does not consider the presence of gradients inside the regions. Therefore we expect the optimal tessellation to trace the isocontours perpendicular to the direction of any gradient that may exist in the signal. The size of the regions will be determined by the intensity of the gradient and the signal-to-noise ratio. Roughly speaking, the size of the isocontours should be of the order of the typical uncertainties $e_{ij\lambda}$ within the region, so that the algorithm can tell that the signal is indeed different from the values in the immediate vicinity.

This expectation is consistent with the behaviour that can be observed in Figures 1 and 2. The area of the synthetic multi-images covered by the ellipse is divided into concentric regions that roughly follow its shape. Most of the tessellations in NGC 2906 tend to feature annuli that trace the isocontours of the considered signal. It is worth noting that, in one of the astrophysical tests (#5 and #6), this refers to the measured intensities of the emission lines, whereas in the other (#7 and #8) it corresponds to the surface brightness at different wavelengths, including not only the emission lines but also the adjacent continuum. In the central part of the galaxy, the signal-to-noise ratio is high. There are many regions consisting of very few spaxels, and all the lines are clearly detected. In the outskirts, where the signal is weaker, the regions are larger, and the overall normalization of the continuum plays a major role in setting their shape.

Since the morphology of the tessellation traces the spatial structure of the input data, the results depend on how these data are fed to the algorithm. The first noticeable difference between the binning modes described in Section 4.3 (and summarized in Table 3) is that the ‘multi- λ ’ procedure yields a single tessellation for all layers. Conversely, the regions of each layer in the ‘monochromatic’ mode have been obtained separately, and hence they are different and independent on one another.

In our synthetic problem, the circle, triangle, and ellipse are, by construction, truly independent objects, and they appear as such when BATMAN is run in the monochromatic mode. In contrast, when the (R,G,B) layers are considered as a multi-image, the overlap regions are *different* from the rest,

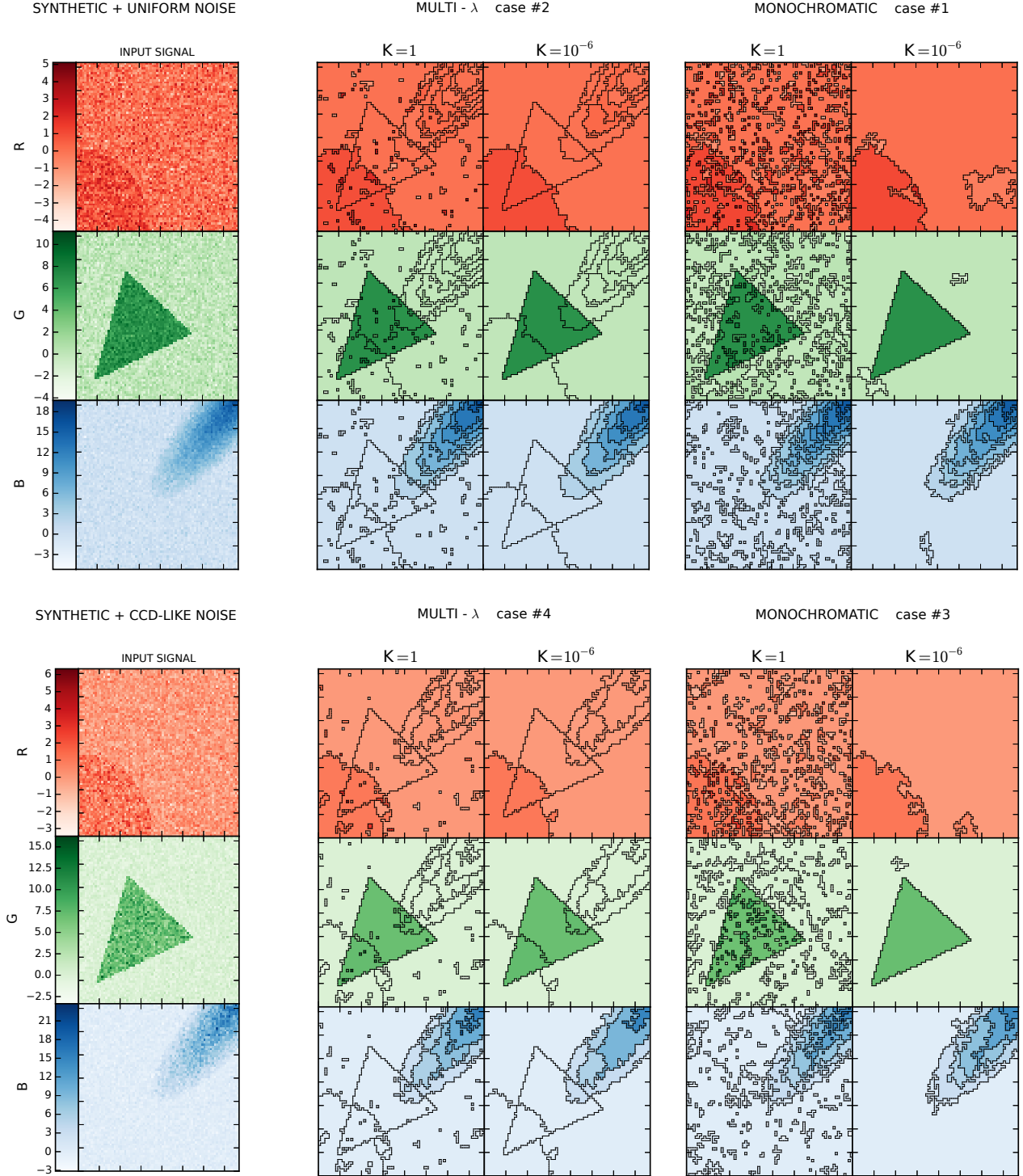


Figure 1. Results of applying BATMAN to our synthetic test cases in ‘multi- λ ’ mode (i.e. considering all the layers in the input data simultaneously; test cases #2 and #4, central columns) and ‘monochromatic’ mode (considering each layer separately, cases #1 and #3, right columns). Two runs are performed in every case, with $K = 1$ (left) and $K = 10^{-6}$ (right). The input signal is displayed in the left column. Rows within each panel correspond to the different layers in the multi-image. The segmentation obtained in every case is indicated with solid black lines and is common to the three layers in the ‘multi- λ ’ test cases #2 and #4 and specific to every one in the ‘monochromatic’ test cases #1 and #3.

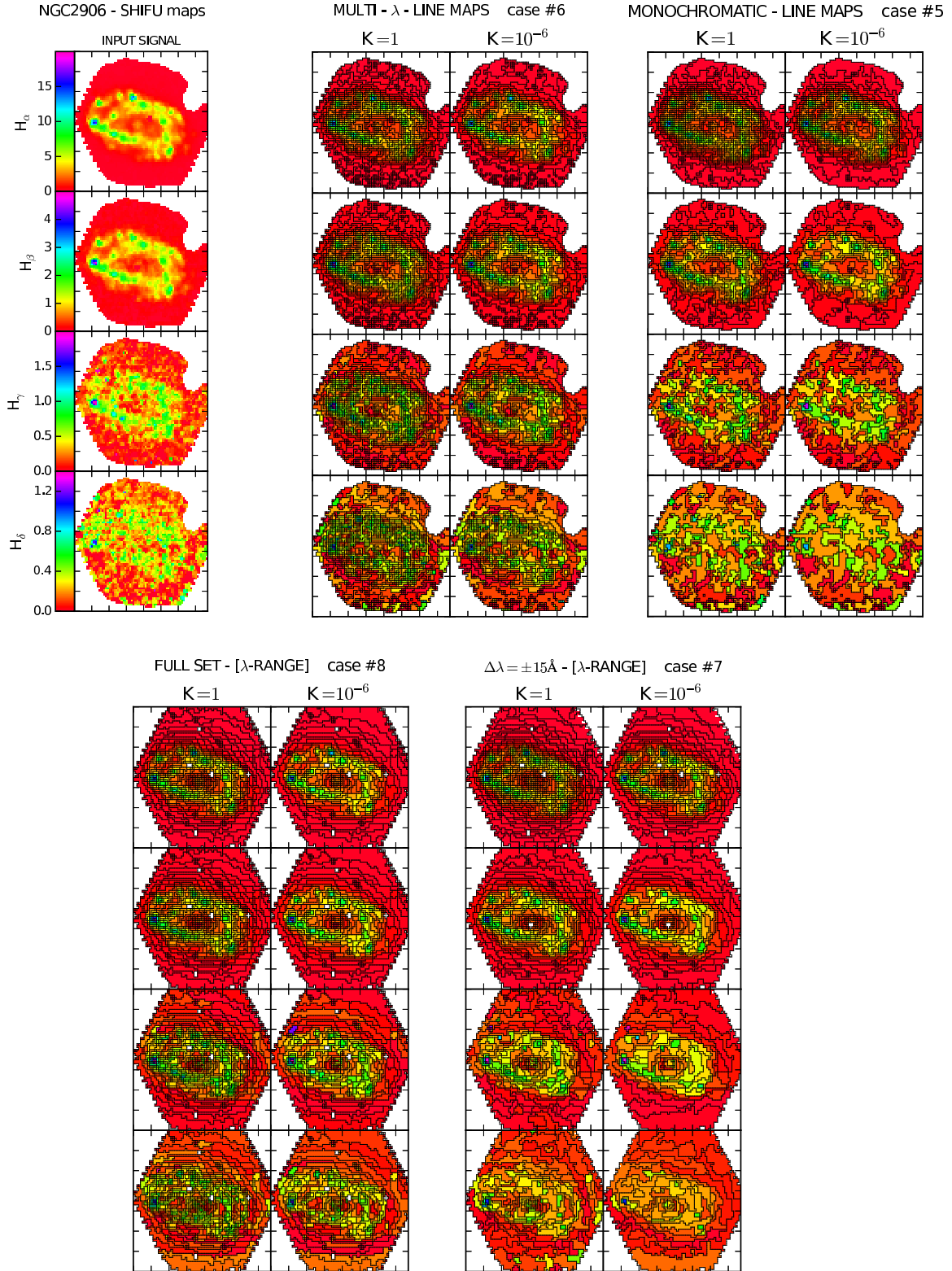


Figure 2. Results for our observational test cases based on CALIFA observations of NGC 2906. The structure of the plot is identical to that of Figure 1 for the upper panels (test cases #5 and #6), which show the output of applying BATMAN to the intensity maps of the Balmer lines measured by SHIFU on a spaxel-by-spaxel basis. The lower panels show the results obtained for test cases #7 and #8, which are conceptually identical to #5 and #6, respectively, but are both multi- λ tessellations of selected slices of the IFS datacube (see the description in Section 4.2 and Table 3); it is impossible to represent the input signal in 2D for these cases. The colour maps denote flux in standard CALIFA units, i.e. $10^{-16} \text{ (erg/s/cm}^2\text{)}$; they correspond to the BATMAN output in test cases #5 and #6 (upper panels), and to SHIFU measurements posterior to the segmentation in test cases #7 and #8 (bottom panels).

as they would correspond to ‘yellow’ and ‘purple’ colours rather than pure ‘red’, ‘green’, or ‘blue’. Hence, BATMAN separates these regions from the ‘main’ objects (circle, triangle, and ellipse) according to the monochromatic definition. Both solutions are equally correct, and which one is to be preferred depends on the specific scientific goals to be achieved. It is entirely up to the user’s judgement to decide what information should be considered relevant in order to decide whether two regions are ‘different’. As an equivalent scientific example, the different layers in a multi-image may correspond to spatially-resolved maps of the *mass* and/or the *fraction* of young stars in a given galaxy, their age, metallicity, radial velocity, velocity dispersion, etc., and one can be interested in analysing all/some of these properties separately (‘monochromatic’ mode) or considering all of them at the same time (‘multi- λ ’ mode). The choice of the input data is a critical step of any analysis, and it will obviously play an important role in determining the solution found by the algorithm.

We also face such a decision in our specific examples of astrophysical test cases, where one may adopt different prescriptions with the ultimate goal of recovering intensity maps of the Balmer emission lines. Is it more appropriate to make a first estimate on a spaxel-by-spaxel basis, and then bin the maps in order to increase the signal-to-noise ratio, or is it better to find an optimal tessellation based on the raw data and then carry out the measurements? In either case, is it more convenient to consider each line separately, or all of them simultaneously? Although it is certainly not our goal to answer these questions, it is interesting to compare several different approaches in order to illustrate the effect that this kind of choices has on the optimal tessellation that BATMAN returns.

When all the SHIFU measurements are taken as input data (case #6, Figure 2), the output tessellation is largely driven by the H α map, which has the highest signal-to-noise ratio. Based on the information provided by this line alone, it is relatively easy to realize that two given spaxels are different if one of them belongs to an HII region and the other corresponds to diffuse emission, where the intensity is considerably lower. The information contained in the weakest lines (H γ and H δ) is also used, but it carries a much lower weight, and it merely represents a minor-order correction. If we run BATMAN separately on each map (case #5, Figure 2), the tessellation obtained for H α is very similar to the multi-wavelength regime. For H β and H γ , the size of the regions increases due to the lower signal-to-noise ratio, but the overall structure is still the same. For H δ , the algorithm adapts to the structures that are present in the input data, but these are severely affected by the noise. This is clearly a disadvantage in our test problem, since we know that the signals in the different layers are not independent. However, this may not be true in the general case (e.g. if there was a signal that was present *only* in H δ), and then it could well be possible that it is more adequate to consider the noisy dimensions separately in order to maximise their weight on the tessellation.

When applied directly to the IFS data (case #7 and #8, Figure 2), the algorithm is able to recover the ring-like structure in all cases, even when only a 30-Å interval around H δ is taken as input. The central regions of the galaxy are always divided into relatively small regions, whereas the outer

parts are tessellated in roughly concentric rings. The overall morphology is actually fairly robust, and the main differences between the different prescriptions are related to the region sizes. In fact, the tessellations obtained from each of the individual intervals (case #7), as well as from the full data set (case #8), are qualitatively similar, at variance with the results based on the synthetic test problems or the SHIFU maps.

5.2 Features on small scales

5.2.1 Random fluctuations

In our synthetic tests (see Figure 1), it is evident that BATMAN recovers many more aggregations of one (or a few) spaxels when only a single ‘layer’ is analysed (‘monochromatic’ mode) than when all layers in the multi-image are considered simultaneously (‘multi- λ ’). These regions are statistically significant according to BATMAN’s criterion, i.e. the posterior probability ratio (19) indicates that it is more likely that the enclosed signal is statistically different from that of the adjacent regions. We do know, however, that these regions may be ‘real’ only in the sense that the input data are indeed different as the result of statistical fluctuations of the noise³. By adding Gaussian noise, some individual spaxels may deviate significantly (e.g. more than three sigma) from the surrounding neighbours, even if the underlying signal is constant. It is in fact relatively likely that two or more (up to a few) adjacent spaxels happen to deviate in the same direction by a significant amount (e.g. -2σ and -2.5σ), and, depending on the adopted prior, BATMAN may classify them as an independent region, statistically different from the surrounding ($\pm 1\sigma$) ‘background’.

As we increase the number of dimensions, the significance of each individual layer decreases. For example, a fluctuation $\Delta = (3.5\sigma_1, -0.5\sigma_2, 1.1\sigma_3, -1.2\sigma_4, 0.2\sigma_5, 0.5\sigma_6)$ in six dimensions yields $\chi^2 = \sum_i \frac{\Delta_i^2}{\sigma_i^2} = 3.5^2 + (-0.5)^2 + 1.1^2 + (-1.2)^2 + 0.2^2 + 0.5^2 = 15.44$, and it has considerably less statistical significance than the 3.5- σ fluctuation in the first layer alone. In terms of χ^2 , it is completely equivalent to a 1.6- σ fluctuation in all layers ($6 \times 1.6^2 = 15.36$). Moreover, if there is indeed a signal over a small region, all spaxels are biased in the same direction, whereas this is very unlikely for a random fluctuation: more precisely, the probability that all n_λ layers in n_{reg} spaxels fluctuate in the same (e.g. *positive* or *negative*) direction by sheer chance decreases as $2^{-n_{\text{reg}} n_\lambda}$, and random aggregations of a few adjacent spaxels are thus much harder to form in the ‘multi- λ ’ case.

For analogous reasons, a true signal is also more difficult to detect: a clear difference in one layer becomes less and less significant if it is diluted among a plethora of other channels carrying unrelated information (and/or noise). In reality, it is far from trivial to discriminate weak signals from random fluctuations, and this is again a challenge left for the final user. In our tests based on astronomical data, the situation is further complicated by the large differences in S/N from

³ This is the so-called *look-elsewhere* effect: a large fluctuation at a random place is extremely unlikely, but a few large fluctuations *somewhere* in the sample are indeed expected.

one line to another as well as the potential presence of observational artefacts. For the H α maps, based on either SHIFU measurements or CALIFA data, including information from the other Balmer lines tends to reduce the number of regions in the areas where the signal is stronger (i.e. within the ring-like structure), preferentially removing regions with one or two spaxels. In the outer parts, though, a number of single-spaxel regions arise in the multi-wavelength tessellations due to the presence of outliers in *any* of the other layers. Although a careful study would be required in order to assess whether these outliers are physical, it is our preliminary impression that most of the single-spaxel regions in the galaxy outskirts are due to artificial features in the data. For all the Balmer lines other than H α , the monochromatic tessellations always yield much larger regions than the multi-wavelength result, because of their much smaller signal-to-noise ratio.

5.2.2 Final number of regions

Regardless of their ultimate physical nature (random fluctuations or real signal), the small-scale structure of the output tessellation is determined by the number, size, and arrangement of the smallest regions that are identified by the algorithm as statistically significant.

As explained in Section 3, the iterative merging procedure continues as long as there is any candidate tessellation $R_{c,n-1}$ where the posterior probability ratio (19) is larger than unity, or, equivalently,

$$\prod_{\lambda=1}^{n_\lambda} (H_\lambda - L_\lambda) \frac{e^{-\frac{(\mu_{A\lambda} - \mu_{B\lambda})^2}{2(\sigma_{A\lambda}^2 + \sigma_{B\lambda}^2)}}}{\sqrt{2\pi(\sigma_{A\lambda}^2 + \sigma_{B\lambda}^2)}} > K \quad (25)$$

Thus, the combined prior parameter K regulates the contrast that is required in order to consider that two regions correspond to different signal values. The effect of considering different values of the combined prior parameter K is illustrated in Figures 1 and 2 by depicting the results obtained for $K = 1$ and $K = 10^{-6}$ in all our example test cases. A decreasing value of K essentially increases the probability of merging regions, and some of the small fluctuations discussed above are removed as a result of a less strict compatibility criterion.

In terms of the prior probability distributions (7) and (15), regulated by the parameters k_λ and k_R , respectively, a low value of K is closer to the ‘non-informative’ (infinite) improper uniform prior, and therefore it means that we have little previous knowledge about the signal and/or that we prefer a segmentation with as few regions as possible. Both statements are absolutely equivalent, both formally as well as in practice. $K \sim 1$ uses the information contained in H_λ and L_λ to minimise the probability that two regions with physically different signal are merged together. Lower values of this parameter reduce the contamination associated to random fluctuations in the input data and increase the signal-to-noise ratio at the expense of missing the weakest genuine signals. The optimal value of K is problem-dependent, and it can only be decided upon trial and error. Fortunately, our results suggest that there are relatively little differences between adopting values as extreme as $K = 1$ and $K = 10^{-6}$, both in synthetic as well as in real data.

To sum up, we can conclude that BATMAN is able to adapt to the structures present in the data, both at large and small scales. The resulting tessellation depends slightly on the adopted priors and, most importantly, on the precise choice of the input data set. Therefore, it is important to devote some time to investigate these issues as a preliminary step of any scientific analysis. The local signal-to-noise ratio and the precise value of the combined prior parameter K (the only free parameter of the algorithm) affect the number and size of the regions identified by the algorithm, especially when gradients are present, but neither of them has a significant effect on the overall morphology of the tessellation. Other aspects, such as the actual shape of the underlying structure, or the statistical properties of the noise, do not seem to play a major role.

5.3 Quality of the reconstruction

One of the reasons to tessellate a multi-image is to obtain a better reconstruction of the underlying signal, especially in the low signal-to-noise regions. In order to calibrate the improvement with respect to the original input data, we study in Figure 3 the distribution of the residuals for our two synthetic test cases, where $S_{ij\lambda}$ is known, in terms of the input errors $e_{ij\lambda}$,

$$\Delta_{ij\lambda} = \frac{\mu_{r_{ij\lambda}} - S_{ij\lambda}}{e_{ij\lambda}} \quad (26)$$

with $\mu_{r_{ij\lambda}} = \mu_{r\lambda}$ for all spaxels $\{i, j\} \in r$, i.e. $r_{ij} = r$. For the input multi-image, where $\mu_{r_{ij\lambda}} = x_{ij\lambda}$, the residuals $\Delta_{ij\lambda}$ follow by construction independent Gaussian distributions with zero mean and unit dispersion. Therefore, the residual maps on the left column of Figure 3 are statistical realizations of white noise for all layers, without any spatial structure, both for the ‘uniform’ and ‘ccd-like’ schemes.

As can be seen in the other maps, the posterior mean values $\mu_{r\lambda}$ obtained for every BATMAN region are representative of the signal values within the region. In the R and G layers, where the signal within the circle, the triangle, and the background is indeed constant, $\mu_{r\lambda}$ is very close to the true value ($\Delta_{ij\lambda} \simeq 0$, white colour in the residual maps) as long as the region is correctly identified.

In some cases, though, a small number of spaxels may be assigned to a region where the underlying signal is actually different: in our synthetic tests, this only happens near the boundaries of the R circle, which are difficult to trace due to the low signal-to-noise ratio ($S/N = 1$). One may expect that such misclassifications yield $\Delta_{ij\lambda} \sim \mathcal{O}(1)$. If the contrast between any two areas is larger than $e_{ij\lambda}$ (as it occurs e.g. in the triangle in the G band), the difference between adjacent spaxels on each side of the boundary is statistically significant. It is thus very unlikely that they become connected by our iterative procedure before merging with the other surrounding spaxels that belong to the correct region. Although these errors are indeed possible, they are restricted to a few isolated spaxels in the G layer of the ‘ccd-like’ case, where $S/N = 3.5$ (see Table 2).

As discussed above, BATMAN can also identify regions around sufficiently high (a few sigma) random fluctuations on small scales when high values of the combined prior parameter K are adopted, especially in the ‘monochromatic’ case. Again, random fluctuations larger than $\Delta_{ij\lambda} \sim \mathcal{O}(1)$

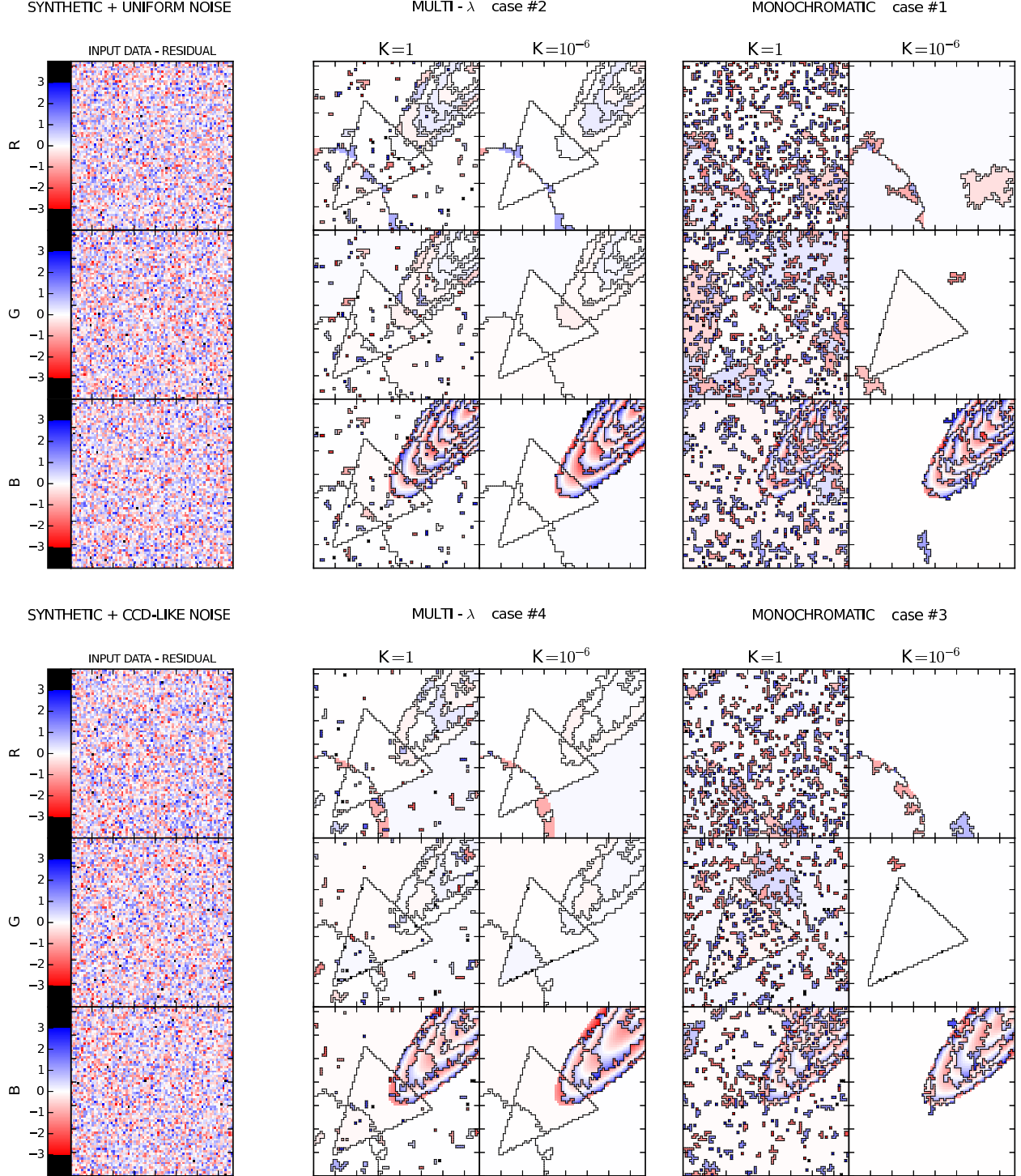


Figure 3. Maps of the residuals $\Delta_{ij\lambda} = \frac{\mu_{r_{ij\lambda}} - S_{ij\lambda}}{e_{ij\lambda}}$ for our synthetic test cases. The structure of the plot is identical to Figure 1.

are statistically very unlikely, and therefore the residuals in the affected spaxels are of this order (red/blue colours in Figure 3).

Errors of similar magnitude also occur in the presence of gradients, where continuous variations of the signal $S_{ij\lambda}$ are physically present in the data, whereas $\mu_{r\lambda}$ is assumed,

by definition, to be constant within every region. Even if the isocontours are correctly traced by the tessellation, and the value of $\mu_{r\lambda}$ is representative of the average signal within the region, we expect, following the same argument applied to the misclassifications, that $\Delta_{ij\lambda} \sim \mathcal{O}(1)$. Otherwise, the deviant spaxel would be associated to the adjacent bin of

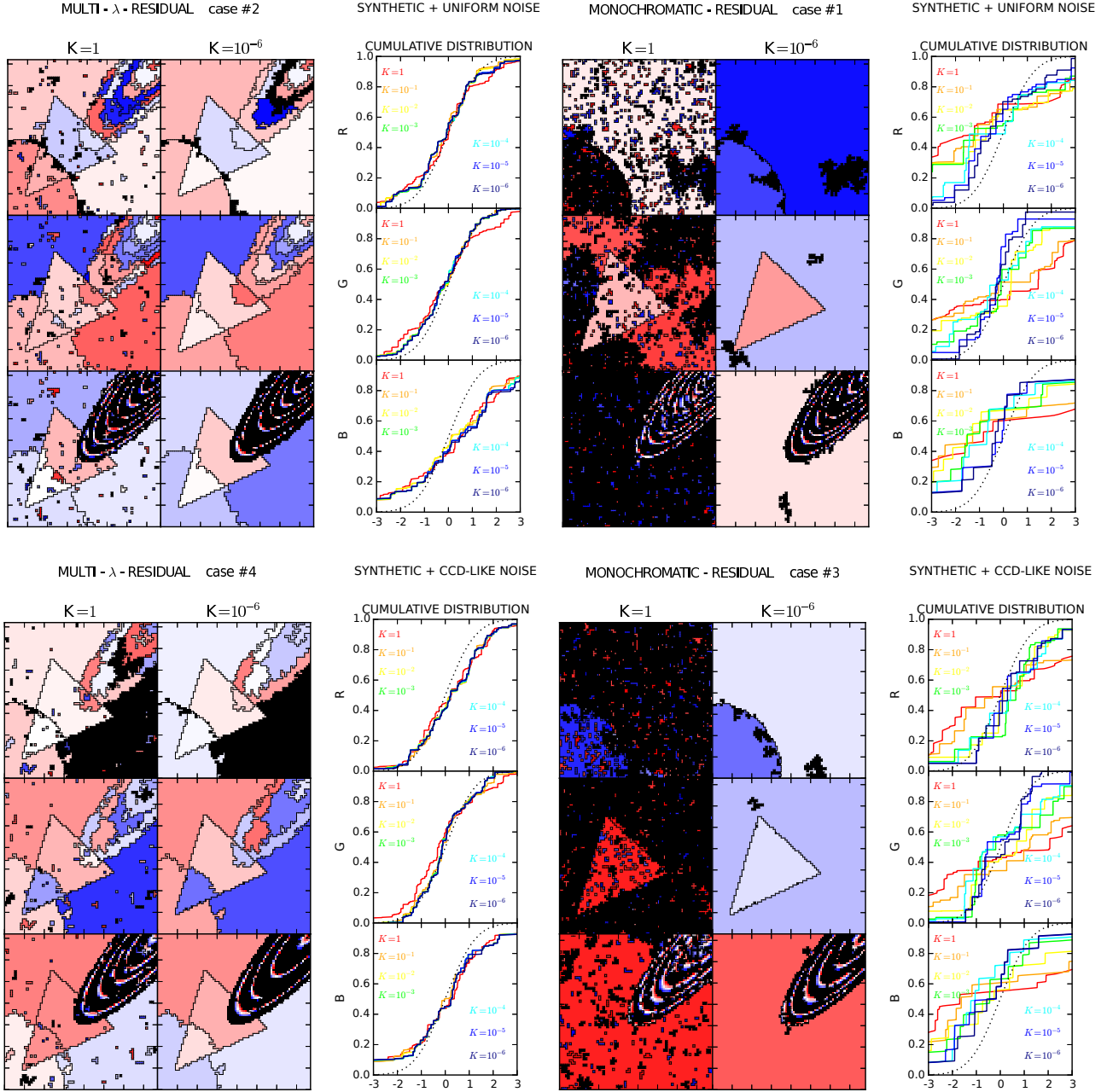


Figure 4. Maps and cumulative distributions of the estimated residuals $\hat{\Delta}_{ij\lambda} = \frac{\mu_{rij\lambda} - S_{ij\lambda}}{\sigma_{rij\lambda}}$ for our synthetic test cases. The structure of the plot is similar to that of Figure 1 except for the missing input data, for which $\hat{\Delta}_{ij\lambda} = \Delta_{ij\lambda}$ (left panels of Figure 3). We include the cumulative distributions of $\hat{\Delta}_{ij\lambda}$ to study the behaviour of the formal errors $\sigma_{r\lambda}$ returned by the algorithm, where different colours correspond to different values of the prior parameter K (see labels in the figure). In order to improve the statistics, the data plotted in these cumulative curves correspond to the results of applying BATMAN to 10 random realizations of the noise. Black dotted curve corresponds to the unbinned input data for the same 10 realizations, where we correctly recover the Gaussian distribution of the noise. Ideally, all the other curves should follow as close as possible this distribution; a more extended range of values of $\hat{\Delta}_{ij\lambda}$ indicates that the formal errors $\sigma_{r\lambda}$ underestimate the true uncertainty (i.e. the rms difference between $\mu_{rij\lambda}$ and $S_{ij\lambda}$).

correspondingly higher or lower intensity. The effect of gradients is illustrated by the ellipse in the B band, where a characteristic red-white-blue pattern is clearly visible within every BATMAN region.

We thus conclude from Figure 3 that the expected values of the posterior probability $\mu_{r\lambda}$ returned by our algorithm provide a more accurate representation of the under-

lying signal than the original input data over most of the multi-image. Only for the ‘catastrophic’ failures discussed above, residuals of the order of $\Delta_{ij\lambda} \sim \mathcal{O}(1)$ may be found.

Another important aspect is the quality of the formal errors $\sigma_{r\lambda}$ derived from equation (12) in Section 2.1. In the ideal case, where all the underlying assumptions are met,

the estimated residual

$$\hat{\Delta}_{ij\lambda} = \frac{\mu_{r_{ij\lambda}} - S_{ij\lambda}}{\sigma_{r_{ij\lambda}}} \quad (27)$$

should follow a normal distribution. However, the presence of gradients in the signal and the identification of spurious regions will cause that the formal errors $\sigma_{r\lambda}$ reported by BATMAN should be treated with caution and merely taken as a (realistic) lower limit to the actual uncertainty in the recovered signal.

A more quantitative assessment is provided in Figure 4, where we plot colour maps of the estimated residual $\hat{\Delta}_{ij\lambda}$. We also present histograms of its cumulative distribution (based on 10 independent realizations of the random noise) for different values of the combined prior parameter K . Using our synthetic tests we find that, for $K < 10^{-3}$ (lines of green-blue colour), the statistical distribution of the estimated residuals $\hat{\Delta}_{ij\lambda}$ is *similar, but not exactly equal* to the expected normal distribution. For $K = 1$, values of $|\hat{\Delta}_{ij\lambda}| > 3$ (black areas in the colour maps) are common in the ‘monochromatic’ case and relatively frequent in the ‘multi- λ ’ analysis. In the latter (and/or for $K < 10^{-3}$), they represent a small fraction of the multi-image and therefore they have a barely noticeable effect on the cumulative distribution.

The most important departure with respect to the ideal behaviour is due to the presence of gradients in the data. Although the algorithm manages to identify regions where the underlying signal is *approximately* constant, equation (12) is based on the assumption of strict equality, and physical variations within the region (of the order of $e_{ij\lambda}$) are not taken into account in the error budget. As can be seen in Figure 4, this source of systematic error may result in $\sigma_{r\lambda}$ underestimating the true residuals in any given spaxel by as much as a factor of the order of two, which arguably constitutes a reasonable upper limit, albeit not necessarily a worst-case scenario.

In order to investigate a more realistic example, let us now try to address the quality of the reconstruction in our astrophysical test cases. As in any practical application, the underlying signal is unknown, and the true solution is not available. In the top panels of Figure 5 we compare the error maps $e_{ij\lambda}$ of the SHIFU spaxel-by-spaxel measurements with the formal errors $\sigma_{r\lambda}$ recovered by BATMAN, both in the ‘monochromatic’ (case #5) and the ‘multi- λ ’ (case #6) binning modes. For test cases #7 and #8, we plot on the bottom panels of Figure 5 the errors $e_{r\lambda}$ reported by SHIFU when the regions defined by BATMAN from the CALIFA data are given as input.

In all cases, we do observe that the formal errors decrease (i.e. the colour maps become redder) and their distribution becomes more uniform thanks to the segmentation. Although the improvement is certainly not surprising in the top panels ($\sigma_{r\lambda}$ decreases roughly as the square root of the number of spaxels in region r), it is an expected but non-trivial result in the bottom panels. In fact, the errors in the line measurements performed by SHIFU after the BATMAN segmentation ($e_{r\lambda}$) are much more similar to the results of the spaxel-by-spaxel analysis ($e_{ij\lambda}$) than suggested by equation (12).

This is due to a combination of different non-linear effects. The most important is arguably the correlation be-

tween the noise in adjacent spaxels in CALIFA data, for which BATMAN assumption of statistical independence is a poor approximation. The propagation of the errors through the SHIFU pipeline takes it into account through a correction factor that attains values of the order of $\sim 2 - 3$ for regions with $\sim 8 - 66$ spaxels (see Appendix A in [García-Benito et al. 2015](#), for details). Other issues, such as potential misclassifications, gradients, and the heterogeneous quality of the observational data (including the error estimates) across different spaxels also contribute to increase the uncertainty.

Yet, the bottom panels in Figure 5 indicate a significant reduction of the errors $e_{r\lambda}$ in the low signal-to-noise areas (e.g. the boundaries) of the multi-image, where large (often annular) regions are identified by BATMAN. This improvement is consistent with the intensity maps displayed in Figure 2, where the level of random fluctuations in these regions is substantially reduced compared to the spaxel-by-spaxel results shown in the top left panels. In areas where the S/N of the CALIFA datacubes is high, our algorithm tends to return regions consisting of one or a few spaxels, and the results (both the intensity and its error) are virtually unchanged with respect to the spaxel-by-spaxel measurements.

We also investigate in Figure 5 the changes in the cumulative distribution of the *estimated* signal-to-noise, comparing $x_{ij\lambda}/e_{ij\lambda}$ (spaxel-by-spaxel measurements) with $\mu_{r\lambda}/\sigma_{r\lambda}$ in the BATMAN output (cases #5 and #6) and with $x_{r\lambda}/e_{r\lambda}$ (SHIFU measurements on BATMAN regions; cases #7 and #8).

First of all, we would like to argue that the interpretation of these quantities is not straightforward, and some of the first impressions conveyed by the histograms in Figure 5 are grossly misleading. In particular, they might seem to suggest at first sight that the best strategy to tackle this particular problem is to run BATMAN on the SHIFU measurements, as this approach leads to the strongest reduction in the formal errors $\sigma_{r\lambda}$ and the largest increase in the estimated S/N (especially for H δ in the ‘monochromatic’ case).

This is certainly at odds with our discussion in Section 5.1, based upon visual inspection of the maps in Figure 2. The original S/N in this particular case is so poor that BATMAN is barely able to even recover the large-scale morphology of the underlying signal. As a consequence, the regions of the tessellation contain spaxels where the signal may vary considerably, and the formal errors are thus expected to be underestimated by about a factor of two, according to our previous test (Figure 4). As mentioned above, the correlation between the noise of adjacent spaxels represents a further source of uncertainty, leading to an additional underestimation of the error that grows logarithmically with the number of spaxels in the region under consideration ([García-Benito et al. 2015](#)).

The improvement of $x_{r\lambda}/e_{r\lambda}$ (cases #7 and #8) with respect to the spaxel-by-spaxel measurements $x_{ij\lambda}/e_{ij\lambda}$ looks much more modest, but we think it is more realistic.

When the signal-to-noise ratio is relatively high (e.g. $S/N > 5 - 10$, inner parts of the galaxy), the CALIFA data are good enough to distinguish fairly minor differences between adjacent spaxels. Since BATMAN’s goal is to keep all the information that is present in the multi-image, the resulting regions tend to be small. Due to the reduced number of spaxels and the presence of correlated noise, the measure-

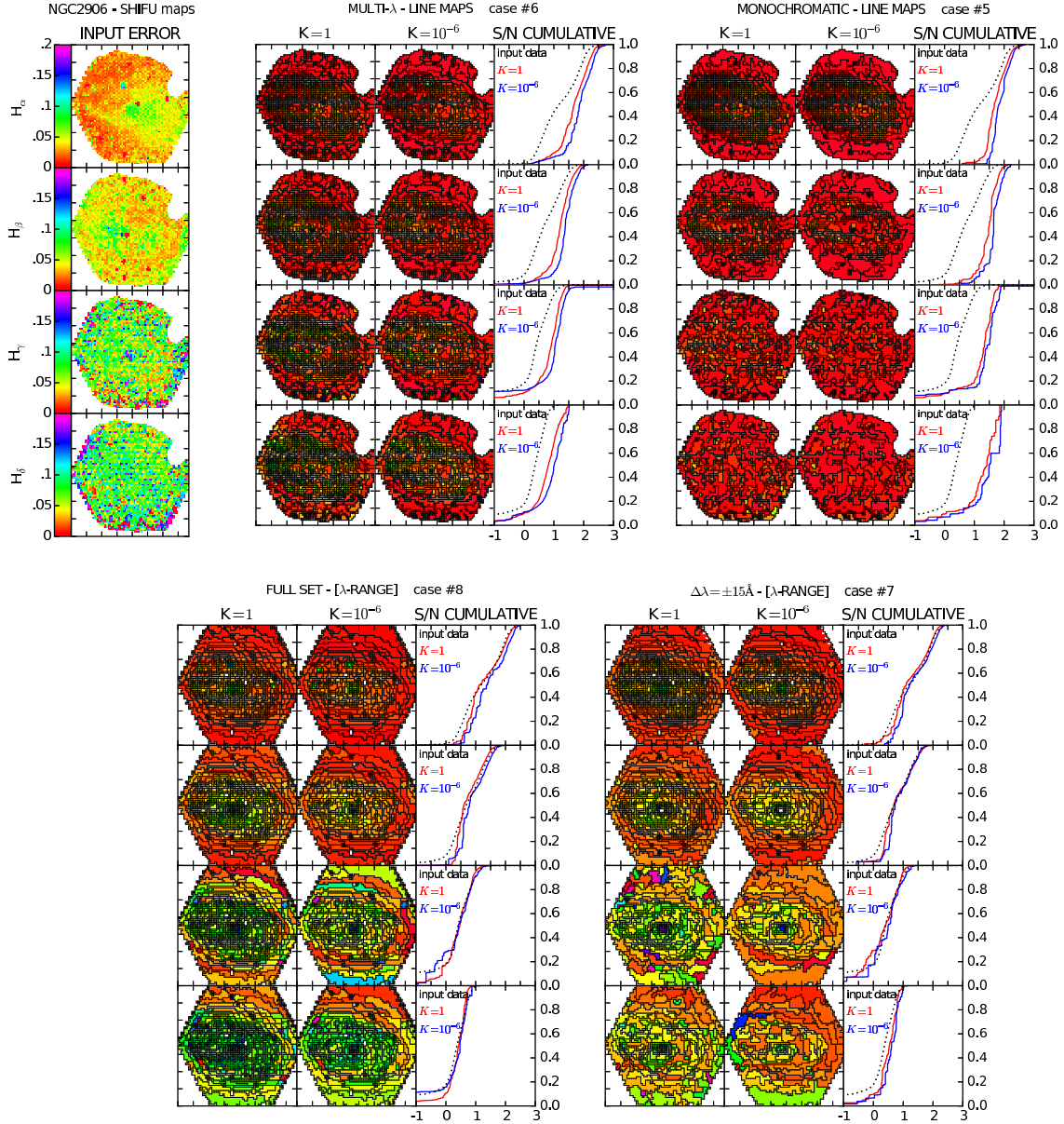


Figure 5. Results of applying BATMAN to our observational test cases, error maps and cumulative distributions of the ‘recovered’ signal-to-noise ratios. The structure of the plot is identical to Figure 2 except for the inclusion of the cumulative distributions. Red and blue cumulative curves correspond to $K = 1$ and $K = 10^{-6}$ results. The dotted black curve corresponds to the SHIFU maps used as input for test cases #5 and #6 and plotted in the left column. The colour maps correspond to flux errors for every line in standard CALIFA units, i.e. $10^{-16}(\text{erg/s/cm}^2)$. The signal-to-noise ratio is represented in logarithmic scale.

ment errors $e_{r\lambda}$ associated to these regions are only slightly smaller than $e_{ij\lambda}$. Whether further merging would be justifiable in order to obtain larger regions and reach a higher signal-to-noise ratio is not a decision to be taken by the algorithm. If we decided that the differences that are present in the data (e.g. the intensity and shape of the stellar continuum) are not relevant for our purpose of measuring the Balmer lines, we should provide a different input dataset (e.g. a continuum-subtracted spectrum) that is free from such irrelevant information.

When $S/N < 3$ (e.g. in the outskirts), Figure 5 shows that BATMAN regions are large enough to decrease the errors, even considering the effect of correlated noise. However,

it is precisely in this regime (particularly for the weakest Balmer lines) that SHIFU tends to fail to identify the emission line and/or return spurious values of both fluxes and errors (e.g. many of the blue and purple dots in the $H\gamma$ and $H\delta$ error maps). These ‘catastrophic’ spaxels make a significant contribution to the fraction of spaxels with $S/N < 1$. If they are not merged with the surroundings (as it is often the case when BATMAN considers the ‘full set’ of spectroscopic data), there is little apparent improvement in the signal-to-noise. If, on the contrary, they are incorporated into the adjacent regions, they may spoil the fit for the whole region, leading to the large blue/purple areas in the ‘ $\pm 15\text{-}\text{\AA}$ ’ panels for $H\gamma$ and $H\delta$.

Moreover, the *estimated* signal-to-noise x/e is a random variable expected to follow a normal distribution centred at the true value of S/N , which has a significant impact on the histograms plotted in Figure 5 whenever the noise is comparable to the signal. In our example, the colour maps in Figures 2 and 5 indicate that both the errors and the amplitude of random fluctuations in the outer regions have decreased by a similar amount when SHIFU is run on the binned spectra, compared to the spaxel-by-spaxel measurements. Although this is certainly a valuable improvement, it barely reflects in the distribution of signal-to-noise ratios.

Although it is not the goal of the present work to deal with all these effects, they illustrate the kind of problems that one may face when analysing real scientific data and the way BATMAN handles them by default. Most importantly, they help showing that the formal errors returned by the algorithm provide an estimate of the true uncertainties roughly as good as the underlying assumptions apply to the particular problem at hand. Thus, we strongly advise not to blindly take them (nor the estimated signal-to-noise ratios) at face value.

6 CONCLUSIONS

This article describes the Bayesian Technique for Multi-image Analysis (BATMAN), a new segmentation algorithm designed to characterize and coherently tessellate simultaneously many layers of a given multi-image, which we define as a dataset containing two regularly-sampled spatial dimensions and an arbitrary number n_λ of ‘spectral’ layers, such as e.g. Integral-Field Spectroscopic (IFS) data.

BATMAN’s tessellation attempts to identify spatially-connected regions that are statistically consistent with the underlying signal being constant, given the information (measurements and corresponding errors) provided in the input dataset. If the difference between any two regions is found to be significant, they are kept separate in order to avoid unnecessary loss of information. It is important to note that these considerations are independent on the signal-to-noise ratio: if two regions carry the same information (have compatible signal within the errors), they should always be merged together; if they do not, it may be completely unphysical to average over them, and BATMAN will keep them separate.

In order to test the performance of the algorithm and provide some guidance to future users, we have created a set of test cases that comprises both synthetic and real data, analysed in different ways. According to the results of these tests, we conclude that:

(i) The output tessellation depends on the precise choice of the input data set, and therefore it is of paramount importance that the user devotes some time to investigate the information that should be considered relevant as a preliminary step of any scientific analysis.

(ii) BATMAN adapts to the spatial structures present in the data for a wide variety of morphologies, regardless of the statistical properties of the noise. By construction, gradients pose a significant challenge to the algorithm. When they are present, the output tessellation tends to trace the isocontour lines.

(iii) The exact number and size of the regions are mainly set by the local signal-to-noise ratio. The higher the S/N of the data, the easiest it is to distinguish whether two spaxels/regions are different. When S/N is low, many spaxels may be consistent with having a similar signal, and only a small number of large-size, independent regions can be identified.

(iv) The precise value adopted for the combined prior parameter K , the only free parameter of the algorithm, also affects the number and size of the regions in the output tessellation by setting the end of the iterative procedure. Lower values of K result in more iterations and therefore a smaller number regions. This may have a substantial impact on the number of (potentially spurious) structures identified on the smallest scales, particularly when $n_\lambda = 1$ (‘monochromatic’ mode).

(v) In the proposed formalism, the expected values $\mu_{r\lambda}$ of the posterior probability distribution of the signal within each region are given by the inverse variance-weighted average (11). Our synthetic tests show that these values provide a good representation of the true signal.

(vi) The formal errors $\sigma_{r\lambda}$, given by expression (12), are indicative of the true uncertainties, but they may underestimate them by as much as a factor of the order of two (e.g. in the presence of gradients and/or spurious regions).

(vii) Our analysis of real astronomical data shows that the segmentation of IFS datacubes is a complex task, involving many issues that are specific to the scientific problem under study. Our tests based on NGC 2906, focusing on the measurement of several Balmer lines, suggest that BATMAN may be most helpful in the low S/N regime. The algorithm is capable of recovering the underlying structure of the object even in the most difficult case (H δ), especially when it is applied directly on the CALIFA data (and not so much when the SHIFU measurements are taken as input). The reduction of the noise with respect to the spaxel-by-spaxel (no binning) approach is clearly visible in the galaxy outskirts.

As a final remark, let us stress once again that, in contrast to many other segmentation algorithms, BATMAN aims to preserve all the spatial information contained in the original data, as long as it is considered statistically significant. Such philosophy represents a new (and, in the opinion of the authors, much needed) approach to the analysis of astronomical images in the advent of the vast amount and spatial resolution of IFS data to come.

ACKNOWLEDGMENTS

Financial support has been provided by research grant AYA2013-47742-C4-3-P from the *Ministerio de Economía y Competitividad* (Mineco, Spain), as well as the exchange schemes ‘Acciones Conjuntas Hispano-Alemanas’ (PPP-Spain-57050803) and ‘Research Grants - Short-Term Grants, 2016’ (57214227) promoted by the *Deutscher Akademischer Austausch Dienst* (DAAD, Germany), and the ‘Study of Emission-Line Galaxies with Integral-Field Spectroscopy’ (SELGIFS, FP7-PEOPLE-2013-IRSES-612701) programme funded by the Research Executive Agency (REA, EU). YA is also supported by contract RyC-2011-09461 of the

Ramón y Cajal programme (Mineco, Spain). RGB acknowledges support from grants AyA2014-57490-P and JA-FQM-2828. This study makes use of the data provided by the Calar Alto Legacy Integral Field Area (CALIFA) survey (<http://califa.caha.es/>), based on observations collected at the Centro Astronómico Hispano Alemán (CAHA) at Calar Alto, operated jointly by the Max-Planck-Institut für Astronomie and the Instituto de Astrofísica de Andalucía (CSIC). Finally, the authors would like to thank the ‘Galaxies and Quasars’ research group at the Leibniz-Institut für Astrophysik Potsdam (AIP), and Peter Weilbacher in particular, for useful discussions and constructive feedback, as well as the anonymous referee for pointing out several aspects (e.g. Table 1, Section 5.2.1, and Appendix B) that have certainly helped to improve the clarity of the paper.

REFERENCES

- Ascasibar Y., Guidi G., Casado J., Scannapieco C., Díaz A. I., 2016, *ArXiv*: 1602.08474,
 Bertin E., Arnouts S., 1996, *A&AS*, **117**, 393
 Cappellari M., Copin Y., 2003, *MNRAS*, **342**, 345
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, **345**, 245
 Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, *MNRAS*, **358**, 363
 Diehl S., Statler T. S., 2006, *MNRAS*, **368**, 497
 Doe S., et al., 2007, in Shaw R. A., Hill F., Bell D. J., eds, *Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI*. p. 543
 Freeman P., Doe S., Siemiginowska A., 2001, in Starck J.-L., Murtagh F. D., eds, *Proc. SPIE Vol. 4477, Astronomical Data Analysis*. pp 76–87 ([arXiv:astro-ph/0108426](https://arxiv.org/abs/astro-ph/0108426)), [doi:10.1117/12.447161](https://doi.org/10.1117/12.447161)
 García-Benito R., Zibetti S., Sánchez S. F., et al. 2015, *A&A*, **576**, A135
 González Delgado R. M., Cerviño M., Martins L. P., Leitherer C., Hauschildt P. H., 2005, *MNRAS*, **357**, 945
 Haffner L. M., et al., 2009, *Reviews of Modern Physics*, **81**, 969
 Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, *MNRAS*, **422**, 1812
 Husemann B., Jahnke K., Sánchez S. F., et al. 2013, *A&A*, **549**, A87
 Koribalski B. S., 2012, *Publ. Astron. Soc. Australia*, **29**, 359
 Krajnović D., Cappellari M., de Zeeuw P. T., Copin Y., 2006, *MNRAS*, **366**, 787
 Makovoz D., Marleau F. R., 2005, *PASP*, **117**, 1113
 Men’shchikov A., André P., Didelon P., Motte F., Hennemann M., Schneider N., 2012, *A&A*, **542**, A81
 Molinari S., Schisano E., Faustini F., Pestalozzi M., di Giorgio A. M., Liu S., 2011, *A&A*, **530**, A133
 Papaderos P., Izotov Y. I., Thuan T. X., Noeske K. G., Fricke K. J., Guseva N. G., Green R. F., 2002, *A&A*, **393**, 461
 Sánchez S. F., Kennicutt R. C., Gil de Paz A., et al. 2012a, *A&A*, **538**, A8
 Sánchez S. F., Rosales-Ortega F. F., Marino R. A., et al. 2012b, *A&A*, **546**, A2
 Sánchez S. F., Pérez E., Sánchez-Blázquez P., et al. 2016, *Rev. Mex. Astron. Astrofis.*, **52**, 171
 Sanders J. S., 2006, *MNRAS*, **371**, 829
 Sanders J. S., Fabian A. C., 2001, *MNRAS*, **325**, 178
 Savage R. S., Oliver S., 2007, *ApJ*, **661**, 1339
 Stetson P. B., 1987, *PASP*, **99**, 191
 Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, **404**, 1639

APPENDIX A: 1D PROBLEM

Let us illustrate the greedy binning procedure that BATMAN performs in order to explore the space of all possible tessellations by means of a very simple problem in one dimension. To this purpose, we have created a 30-element vector with values 0, 4, and 2 in the ranges [0,9], [10,19] and [20,29], respectively (blue line in top left panel of Figure A1). We have added Gaussian noise with a dispersion $\sigma = 1$ (blue shaded region) to the ‘noiseless’ ‘1D-image’, as in the synthetic test cases of Section 4.1, to obtain an input vector (signal and noise) for our binning algorithm (black solid line).

We apply BATMAN to this problem and allow the algorithm to continue binning even when it is no longer ‘convenient’ according to our Bayesian criterion (see Section 3), i.e. when the ratio in equation (19) is smaller than unity for any new tessellation obtained by merging two adjacent regions, the evidence for any such model decreases with respect to the current solution, and BATMAN considers that any further co-addition of regions leads to a loss of statistically-significant information.

The entire process is presented in Figure A1, where the top row shows (left) the input (black) and true (blue) signals, as well as the variation of the evidence ratio (middle) and the evidence (right) for the current tessellation $\mathcal{E}(R)$ according to expression (13) as a function of the number of iterations (or, equivalently, the number of regions in the tessellation). The horizontal solid red line (ratio equal to one) in the middle top panel corresponds to the stopping criterion of our algorithm, and the iteration where it is reached is indicated by a vertical dotted red line on both the middle and right panels. We have verified that this iteration is always close to (and very often coincides) with the absolute maximum of the evidence, marked by the solid green line in the right panel (see Appendix B).

The rest of the panels (second to seventh rows) show the iterative merging of vector elements into larger one-dimensional regions. In addition to the true signal (blue) and input measurements (black) we have plotted the recovered signal (red solid line) and estimated error (red shaded area) at every iteration. BATMAN starts by merging the most similar vector elements, which makes iterations 1 (connecting elements 16 and 17) and 2 (elements 10 and 11) almost unnoticeable (the underlying black curve is barely visible). However, one can see that the noise (red shade) is reduced (see iteration 2) even when the change in the recovered signal is very small. The binning process continues, and the red line progressively resembles the blue one, while the red shade tightens around it (the noise is reduced). Iteration 27 shows the most likely tessellation (last one with a positive logarithm of the evidence ratio), with iteration 28 (last panel) containing only two regions being clearly less representative of the input signal, not only in mathematical terms but also from visual inspection.

APPENDIX B: STOPPING CRITERION

In principle, there is no guarantee that our greedy procedure does not converge to a local maximum of the evidence rather

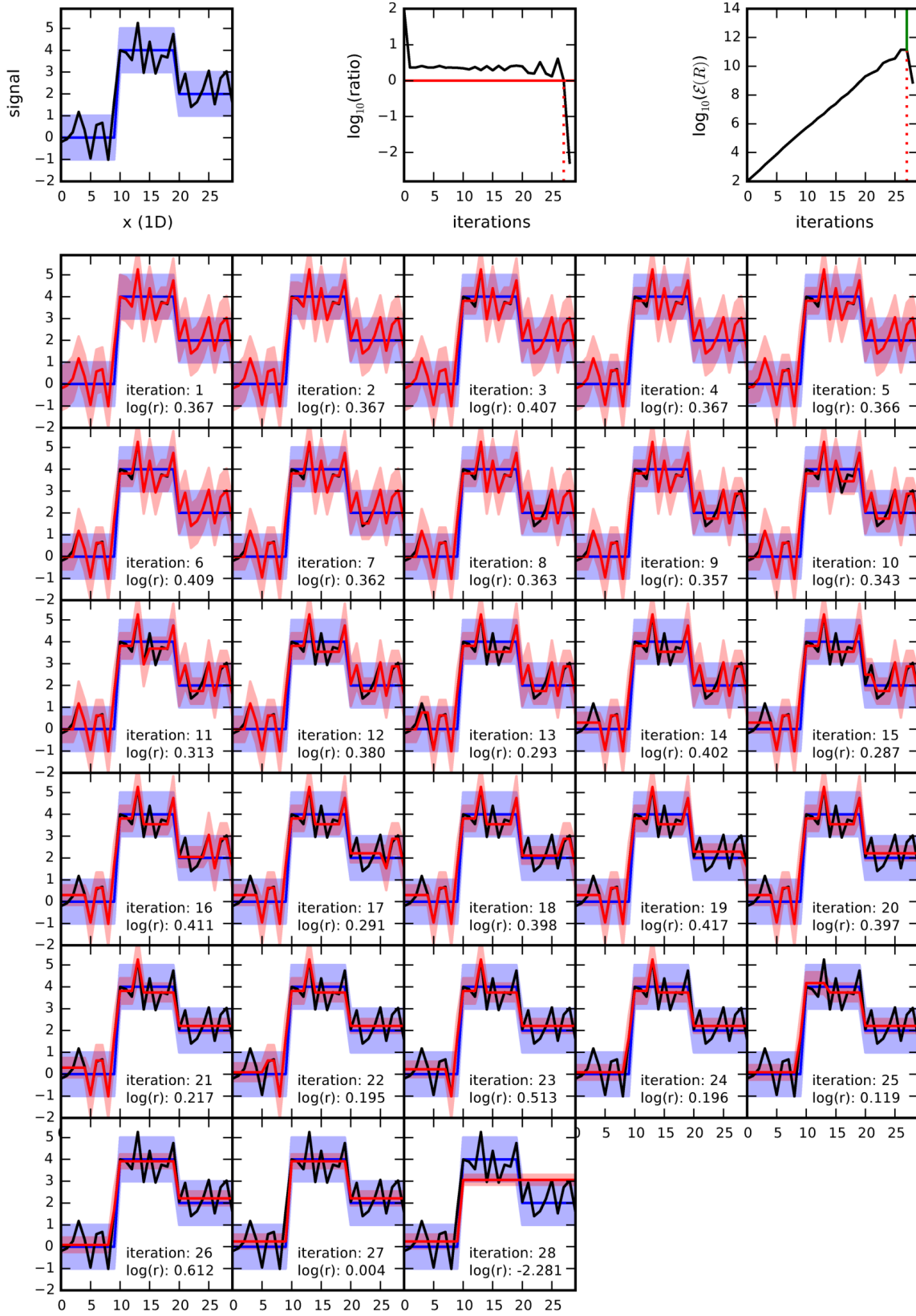


Figure A1. Illustration of the binning procedure in a 1D problem. Top left panel shows the true signal (blue line), $1 - \sigma$ error (blue shaded region), and a random realization of the measurements (black line). The value of the ratio (19) for the candidate tessellation selected at every iteration and the evidence $\mathcal{E}(R)$ are depicted in the top middle and top right panels, respectively. The horizontal red solid line on the middle panel indicates the stopping condition (ratio smaller than unity), while the vertical dotted line marks the iteration at which it is attained. The top right panel shows that it coincides with the maximum value of the evidence (green solid line). The rest of the panels plot the recovered signal μ and the estimated error σ as a red solid line and red shaded region, respectively. The logarithm of the ratio (positive until iteration 28) is quoted on every panel.

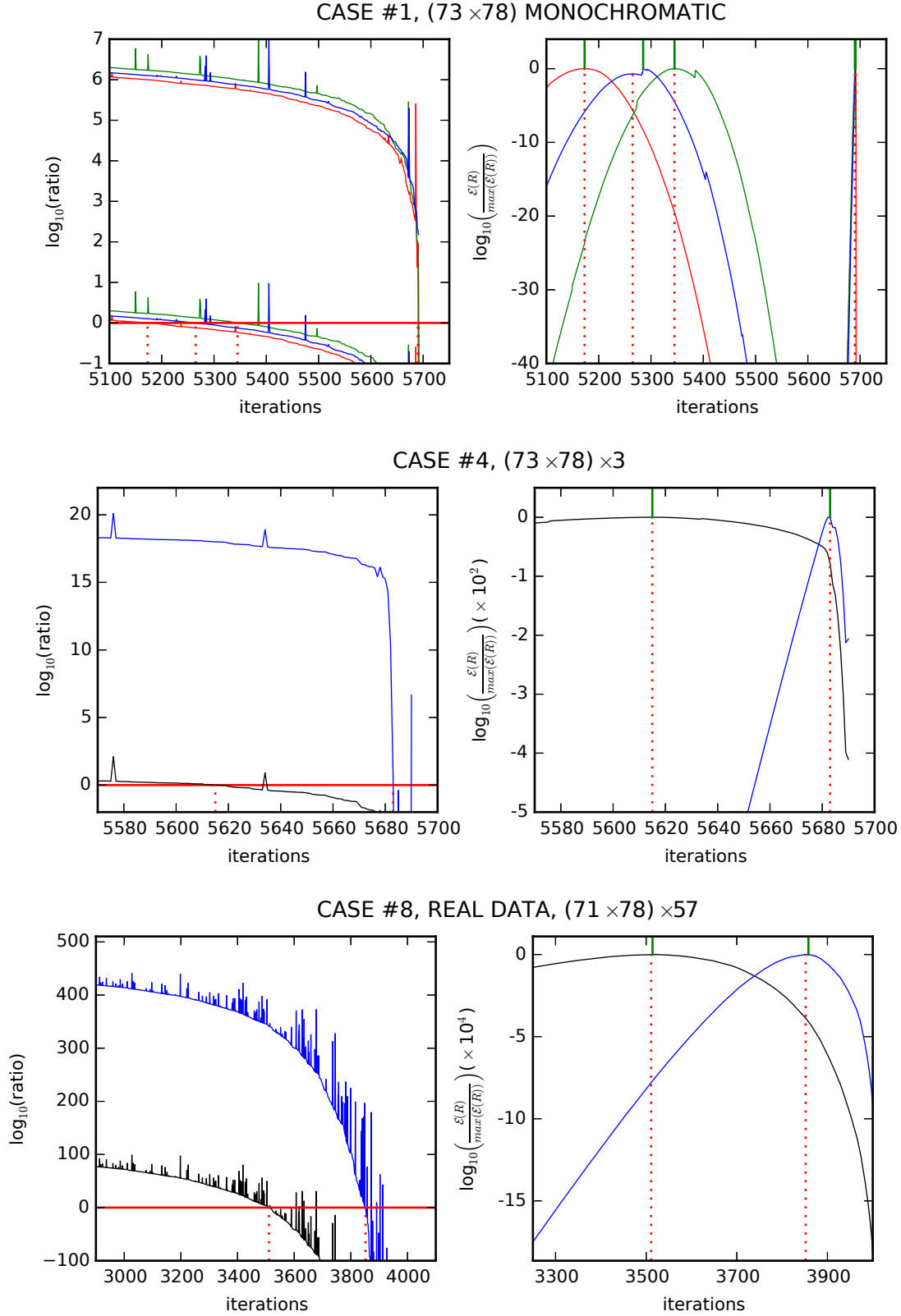


Figure B1. Illustration of the stopping criterion (red solid and dotted lines) in relation to the ratio (left) and the evidence (normalized for comparison, right) at every iteration for the test cases #1 (top), #4 (middle), and #8 (bottom). Top panels present 6 curves corresponding to the two different values of the K parameter (1 lower curves and 10^{-6} higher ones) for the different RGB images. Middle and bottom panels show two curves corresponding to the different values of the parameter K (1 in black and 10^{-6} in blue). Red solid line (left panels) represents the threshold used as stopping condition. Red dotted line (both panels) marks the stopping iteration. Green solid line (right panels) corresponds to the maximum of the evidence. Values for the stopping and evidence maximum iterations are depicted in Table B1.

test case	K	stopping it.	$\max(\mathcal{E}(R))$ it.
# 1 - R	1	5173	5173
	10^{-6}	5691	5691
# 1 - G	1	5345	5345
	10^{-6}	5690	5690
# 1 - B	1	5265	5285
	10^{-6}	5691	5691
# 4	1	5615	5615
	10^{-6}	5683	5683
# 8	1	3511	3514
	10^{-6}	3852	3858

Table B1. Values of the stopping (third column) and maximum evidence (fourth column) iterations, corresponding to the red dotted and green solid lines in Figure B1.

than finding the global optimum. While more elaborate (e.g. Monte Carlo) techniques would allow a more exhaustive exploration of the full tessellation space, here we present the results of a simpler test to show that the adopted stopping criterion, i.e. evidence ratio (19) smaller than unity, yields a nearly optimal segmentation, even when it does not always coincide with the absolute global maximum of the Bayesian evidence $\mathcal{E}(R)$.

As we did for the one-dimensional problem presented in Appendix A, we have re-run BATMAN for our test cases #1 (monochromatic mode on R, G, and B images with uniform noise, i.e. 3 times), #4 (multi- λ mode, RGB image with CCD-like noise) and #8 (real data, full set, directly binning the IFS datacube) allowing the algorithm to proceed beyond the adopted stopping condition. Figure B1 shows the ratio (left column) and the evidence (right column) as a function of iteration number for each of these test cases. Once again, the red solid horizontal line and the red dotted vertical line indicates BATMAN's stopping criterion. In most cases, the evidence does not increase further if we allow for additional merging of regions (as shown in Figure A1 for the one-dimensional test problem discussed in Appendix A), and the stopping condition coincides almost exactly with the maximum evidence, marked by a vertical green line on the right column.

These results suggest that our condition based on the evidence ratio provides a valid (and efficient) criterion to select the final tessellation. Even in those cases where the stopping iteration does not coincide with the maximum of the evidence (e.g. cases #1 and # 8), the values are reasonably close (see Table B1). Nonetheless, we would like to point out once again that, as can be readily seen in Figure B1, the Bayesian evidence does depend significantly on the adopted value of the combined prior parameter and the chosen input data set. These (problem-dependent) decisions, especially the latter, are much more relevant than the stopping criterion in determining the optimal segmentation returned by the algorithm. The importance of experimenting by trial and error in each particular case can hardly be overstressed.